

# Kieliympäristön vaikutus S2-oppijoiden sanaston laajuuteen

Auguste Gronskeyte

Pro gradu -tutkielma

Turun yliopisto

Kieli- ja käännöstieteiden laitos

Suomen ja sen sukukielten maisteriohjelma, suomen kieli

Marraskuu 2020

*Turun yliopiston laatu järjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.*

TURUN YLIOPISTO

Kieli- ja käännöstieteiden laitos / Humanistinen tiedekunta

GRONSKYTE, AUGUSTE: Kieliympäristön vaikutus S2-oppijoiden sanaston laajuuteen

Pro gradu -tutkielma, 50 s., 5 liites.

Suomen ja sen sukukielten maisteriohjelma, suomen kieli

Marraskuu 2020

---

Suomi toisena kielenä opiskelijoiden suomen kielen taitoa ja kielitaidon riittävyttä töissä pärjäämiselle tutkivien tutkimusten määrä on kasvussa S2-tutkimuksen alalla. Yleisesti ajatellaan, että maahanmuuttajatyöntekijöillä tulisi olla tarpeeksi laaja ja työn suorittamista varten riittävä kielitaito ja sanavarasto, eikä huomioda, että kielitaitoa voitaisiin edistää työympäristössä. Oleskelu kohdemaassa ilman aktiivista päivittäistä kielenkäyttöä eri arkitilanteissa ei tue kielen kehittymistä. Ekologisesta näkökulmasta ympäristö tarjoaa suomen kielen oppimisen mahdollisuuksia, mutta tarvitaan myös henkilön omaa aktiivisuutta ja kykyä huomata tarjotut affordanssit. Formaali opetus ei tarjoa tarpeeksi aitoja mahdollisuuksia käyttää suomen kieltä, mutta oppija itse voi harjoitella kieltä informaaleissa ympäristöissä, kuten aktiviteettien kautta vapaa-aikanaan. Myös työssäkäynti tarjoaa mahdollisuuksia laajentaa sanastoa. Tämän tutkimuksen tavoitteena on selvittää, miten kohdekielen kieliympäristö vaikuttaa S2-oppijan sanaston laajuuteen.

Aineistona käytän Yleisten kielitutkintojen suomen keskitason testiaineistoa, jonka Jyväskylän yliopiston Soveltavan kielitutkimuksen keskus on kerännyt ennen vuotta 2011 Yleisten kielitutkintojen korpukseen. Kohderyhmäksi valitsin venäjänkieliset informantit. Tarkastelen heidän kirjoittamiaan 80 mielipidetekstiä. Aineisto koostuu kahdesta ryhmästä. Ensimmäinen ryhmä sisältää ne informantit, jotka altistuvat kielelle päivittäin kotonaan, työssään, vapaa-aikanaan ja opiskelussa. Toinen ryhmä koostuu informanteista, joilla on rajalliset suomen kielen käyttömahdollisuudet. He käyttävät kieltä jonkun verran vapaa-aikanaan ja opiskelussa. Tutkimuksessani kiinnitän eniten huomiota työympäristön vaikutukseen sanastoon. Analyysiosassa metodina käytän leksikaalisen diversiteetin laskemista MTLN-mittarin avulla, ja vertailen molempien ryhmien MTLN-arvoja. MTLN-mittari mittaa sanaston monimuotoisuutta. Tilastollisena menetelmänä käytän monimuuttajaista lineaarista regressiota.

Tutkimukseni tulokset osoittavat, että työllisyystilanne vaikuttaa leksikaaliseen diversiteettiin. Ensimmäisen ja toisen ryhmän MTLN-arvojen vertailu osoitti, että työssäkäyvillä on korkeammat MTLN-arvot. Tutkimuksessa kävi myös ilmi, että koulutuksella ja työllistymistilanteella on yhteys sanaston laajuuteen. Korkeakoulutuksella on vaikutus sanaston laajuuteen, jos henkilö on työtön. Korkeakoulutetut, jotka eivät käy töissä, hallitsevat laajemmin sanastoa verrattuna niihin, jotka eivät ole korkeakoulutettuja eivätkä käy töissä. Rakentamani lineaarinen regressiomalli vahvistaa, että työllistyminen vaikuttaa sanaston laajuuteen.

Asiasanat: leksikaalinen diversiteetti, sanasto, kieliympäristö, altistuminen kielelle, suomi toisena kielenä

## Sisällys

1	Johdanto .....	2
2	Sanaston oppiminen osana toisen kielen oppimista.....	6
2.1	Ekologinen lähestymistapa kielen oppimiseen .....	6
2.2	Sanaston oppimiseen vaikuttavat tekijät .....	9
2.2.1	Sanaston tutkimiseen liittyvä terminologia.....	12
2.2.2	Sanaston rooli eurooppalaisessa viitekehyksessä .....	13
2.3	Oppimisympäristö ja kielen oppiminen .....	15
2.3.1	Oppimiskontekstien tyypit ja oppimismuodot .....	15
2.3.2	Oppimiskontekstin rooli sanaston oppimisessa opiskelu ulkomailla konteksti .....	18
2.3.3	Oppimisympäristö formaalin oppimisen ulkopuolella.....	19
3	Aineisto ja menetelmät.....	21
3.1	Yleisten kielitutkintojen kuvaus.....	21
3.2	YKI-korpus .....	22
3.3	Oma aineisto .....	23
3.4	Leksikaalinen diversiteetti .....	25
3.5	MTLD leksikaalisen diversiteetin mittarina .....	27
3.6	Ohjelmat.....	28
4	Aineiston analyysi.....	29
4.1	Muuttujat.....	29
4.2	Aineiston MTLD-arvojen vertailu .....	33
4.3	Lineaarinen regressioanalyysi.....	39
5	Yhteenveto .....	43
5.1	Johtopäätökset.....	43
5.2	Pohdinta .....	45
	Lähteet.....	46
	Aineslähteet.....	46
	Muut lähteet .....	46
	Liite 1 Taustatietolomake.....	51
	Liite 2 Taitotasokuvaukset .....	55

# 1 Johdanto

Kahden viime vuosikymmenen aikana maahanmuutto Suomeen on kasvanut. Tilastojen mukaan vuonna 2019 Suomessa asui 423 494 ulkomaalaistaustaista henkilöä, mikä on noin 8 % väestöstä. Tilastokeskuksen tietojen mukaan Suomen kansalaisuuksia myönnetään joka vuosi noin 8 000–12 000. Vuonna 2019 Suomen kansalaisuus myönnettiin 9 649 henkilölle. 2010-luvulla eniten Suomen kansalaisuuksia ovat saaneet Venäjän, Somalian, Irakin ja Viron kansalaiset. Vuonna 2019 Suomessa asui 412 700 vieraskielistä, eli henkilöitä, joiden äidinkieli on jokin muu kuin suomi, ruotsi tai saame. Se on 7,5 prosenttia koko väestöstä. Suurin ulkomaalaistaustaisten lähtömaa on entinen Neuvostoliitto. Venäjää äidinkielenään Suomessa puhuu lähes 82 000 henkilöä. Vuonna 2004 kansalaisuuslaki (359/2003) uudistettiin, mikä loi mahdollisuuden saada kaksoiskansalaisuuden. Vuonna 2012 tuli voimaan kansalaisuuslain (579/2011) muutos, jossa vaadittava asumisaika lyhennettiin kuudesta vuodesta viiteen. (Tilastokeskus.)

Kun maahanmuuttajat muuttavat Suomeen, he eivät pysty heti aloittamaan töitä, johon heillä on pätevyys kotimaassaan, koska heidän täytyy saada pätevyys työskennellä kohdeammattissaan Suomessa. Suomessa säännellyissä ammateissa, kuten opetusalan, varhaiskasvatuksen, terveydenhuollon, sosiaalihuollon, liikenteen, pelastusalan, eläimiin ja elintarviketurvallisuuteen liittyvissä tehtävissä vaaditaan viranomaisen päätös tutkinnon tunnustamisesta, jos kyseessä oleva tutkinto on suoritettu ulkomailla. Tutkinnon rinnastamisen lisäksi pitäisi osoittaa, että henkilö osaa tarpeeksi hyvin suomen kieltä. Yksi tapa osoittaa kielitaito on Yleisen kielitutkinnon (YKI) suorittaminen. Viime aikoina maahanmuuttajalääkäreiden ja -sairaanhoitajien suomen kielitaidon riittävyys ja vaatimukset ovat olleet suomenkielisten tutkimusten mielenkiinnon kohteena. Kela ja Komppa (2011) ovat tutkineet maahanmuuttajasairaanhoitajien työkieltä ja Seilonen ja Suni (2016) ammatillisen suomen kielen taitoa. Maahanmuuttajalääkärien suomen kielen taitoa ja lääkärin työtehtävien kielellistä vaativuutta on tutkinut Tervola (2019). Suomessa englannin kielellä sairaanhoitajaksi opiskelevia vaaditaan hallitsemaan riittävä suomen kielen taito jo ennen kuin he aloittavat varsinaista työtä. Tästä herää kysymys, miten opiskelija voi oppia ammattikielen sanastoa ilman, että osallistuu työssä tapahtuvaan vuorovaikutukseen. Kielitaidon kehitys tapahtuu juuri työympäristössä, eikä kielitaito voi olla valmis ennen työn aloittamista. Sen takia työympäristön pitäisi toimia myös kielenoppimisympäristönä. (Virtanen 2011, 2017a, 2017b.)

Tämän tutkimuksen tavoitteena on arvioida, miten suomenkielinen ympäristö vaikuttaa suomen kielen oppijan sanavaraston laajuuteen. Tutkin sanaston laajuutta tässä tutkimuksessa leksikaalisen

diversiteetin avulla. Oletan, että asuminen kohdekielisessä maassa vaikuttaa opittavan kielen taitoihin, ja voi jopa nopeuttaa sen oppimista. Tutkittava ryhmä on äidinkielenään venäjää puhuvat YKI-tutkinnon keskitason suorittajat. Tutkimuksessa minua kiinnostavat ne osallistujat, jotka käyttävät suomen kieltä töissä, eli käyttävät kieltä päivittäin ja ne, jotka eivät käy töissä. Vertaan heidän suomen kielen sanaston osaamistaan kirjoitustehtävissä keskenään. Tutkimuskysymykseni on seuraava:

*Miten työssä käyminen ja korkeakoulutus vaikuttavat sanaston laajuuteen?*

Olen jakanut informantit kahteen ryhmään, jossa ryhmien jäsenet on valittu sillä tavalla, että ryhmistä ilmenee kielenkäyttötilanteet. Hypoteesini on, että henkilöiden, jotka käyvät töissä, sanavarasto on laajempi kuin niillä, jotka ovat työttömiä. Oletan, että kohdekielen kieliympäristö – vapaa-ajan aktiviteetit, kuten harrastukset, television, radion ja sosiaalisen median seuraaminen sekä sosiaaliset kontaktit, esimerkiksi suomalainen puoliso, ystävät ja tutut, joiden kanssa saa harjoitella suomen kieltä – voivat positiivisesti vaikuttaa sanaston laajuuteen. Olen jakanut nämä ryhmät myös koulutuksen mukaan, koska tämän tutkimuksen kannalta on tärkeää saada tietoa, miten pohjakoulutus vaikuttaa sanaston laajuuteen. Toinen hypoteesini on se, että korkeakoulutettujen sanasto on laajempi kuin ei-korkeakoulutettujen. Mielestäni on tärkeää ottaa koulutus mukaan tähän tutkimukseen, koska informanttien kielen oppimisen polut voivat olla erilaisia. Jokaisella oppijalla on oma koulutustaustansa, esimerkiksi suorittiko hän peruskoulun, korkeakoulun vai yliopiston, kuinka kauan hän on opiskellut kohdekieltä, tai opiskeliko hän kurssilla vai itsenäisesti. Suomeen muuton syyllä voi myös olla merkittävä rooli. Esimerkiksi avioliitto suomalaisen kanssa voi tarjota mahdollisuuden käyttää kieltä päivittäin kotona. Omassa tutkimuksessani oppimiskontekstin perusteella en voi luokitella osallistujia selkeihin ryhmiin.

Kun puhutaan maahanmuuttajista ja heidän työllistymisestään, pääongelma työllistymisen vähäisyydessä on puutteellinen suomen kielen taito. Kuitenkaan ei ole itsestään selvää, millainen on riittävä kielitaito töissä pärjäämiseen. Käsite ”kielitaidon riittävyys” ei ole niin yksikertainen, vaan se riippuu siitä, mistä näkökulmasta kielitaitoa katsotaan. (Pöyhönen & Tarnanen 2011: 143.) Nyt suomalaisessa sosiaalisessa mediassa puhutaan siitä, että kotouttaminen Suomessa ei ole onnistunut, ja että suurena ongelmana on heikko suomen kielen taito (Kauppalehti 2018). Ongelma on se, että luokahuoneopetuksessa opettaja ei pysty tarjoamaan aitoja vuorovaikutustilanteita tai riittävän monipuolisia tilanteita, jotka myös vastaisivat mahdollisimman hyvin jokaisen oppilaan tarpeita (Komppa 2015). Sekä suomi toisena kielenä -alan opetus että tutkimus ovat jättäneet kieliympäristön vähälle huomiolle. Suni (2008) huomaa, että ”<...> toisen kielen pedagogiikan tulisi olla ennen

kaikkea spontaaniin oppimiseen reagoivaa ja sitä jäsentävää ja tukevaa eikä niinkään ulkomaailman viestintätilanteisiin valmentavaa.”

Riittävää toisen kielen sanaston osaamista tarvitaan, jotta kommunikaatio kohdekielen kielenkäyttäjän kanssa sujuu, ja jotta varmistutaan siitä, että molemmat osapuolet ymmärtävät toisiaan. Viestin välittämiseen ei riitä kieliopin osaaminen, vaan tarvitaan sanoja ja niiden monimerkityksellisyyksien hallitsemista. Asumista kohdekielen ympäristössä usein pidetään riittävänä edellytyksenä toisen kielen oppimista varten, mutta henkilö ei altistu kielelle pelkästään asumalla kohdemaassa. Voi olla, että kielenoppijalla ei ole mahdollisuuksia luoda sosiaalisia kontakteja esimerkiksi työn kautta, koska töitä ei ole, tai työkieli on kansainvälisenä työkielenä pidetty englanti, jolloin mahdollisuudet osallistua sosiaalisiin tilanteisiin ovat vähäisiä. Altistumiselle tarvitaan aktiivista kohdekielen käyttämistä, eli aitoa vuorovaikutusta kohdekielen puhujien kanssa. Sen takia ympäristöä ei voi pitää itsestään selvänä keinona kohdekielen sanaston oppimiseen, koska ensin pitäisi löytää pääsy kielenkäyttäjyhteisöön. (Honko 2015.) Ekologisesta näkökulmasta oppijalta vaaditaan oma-aloitteista toimijuutta, koska passiivinen oppija ei opi kieltä. Myös kohdekielen syötöksen määrä ja laatu ovat tärkeitä. Kieliympäristö tarjoaa kielenkäyttömahdollisuuksia, ja oppijan tehtävä on poimia ne kielelliset seikat, joita hän tarvitsee tai joista on hyötyä. Oppija ei omaksu kaikkea ympäristöstä. (van Lier 2000, 2007.)

Tämän tutkimuksen aineisto on kerätty Yleisten kielitutkintojen korpuksesta, jonka on kerännyt Jyväskylän yliopiston Soveltavan kielentutkimuksen keskus (YKI-korpus). Korpus sisältää Yleisten kielitutkintojen testiaineistoa. YKI-aineistosta on tehty tutkinnon kahden vuosikymmenen olemassaolon aikana runsaasti tutkimuksia, joissa on tutkittu osallistujien taustoja ja tiettyjä äidinkieliä, kuten viroa ja venäjää. Tuorein tutkimus, jossa kohderyhmänä oli venäjänkieliset, oli Suomen Akatemian rahoittama hanke *Integroituminen ammattiyhteisöihin Suomessa* (2011–2014). Hankkeen tutkimuksen aiheena oli maahanmuuttajien integroituminen työelämään. Yksi hankkeen tavoitteista oli arvioida, millainen rooli kielitaidolla on työikäisten maahanmuuttajien integroitumisessa työ- ja ammattiyhteisöihin Suomessa. Tutkimuksen tulokset osoittivat, että suomen kielen taidolla ja integroitumisella työyhteisöihin ei ole suoraa yhteyttä. IT-alan työtehtävissä ja akateemisissa työyhteisöissä työkielenä toimii englanti. Kuitenkin suomen ja venäjän kielten osaaminen voivat auttaa löytämään töitä Suomessa ja nopeuttaa integroitumista suomalaiseen työyhteisöön. (Rynkänen 2015.) Lähempänä tutkimukseni aihetta on *Venäjää äidinkielenään puhuvien maahanmuuttajien suomen kielen taito yleisten kielitutkintojen valossa* -tutkimus (Ahola & Lammervo 2013), jossa YKI-aineiston pohjalta on tutkittu venäjänkielisten osallistujien testituloksia ja taustatietoja: ikää, sukupuolta, koulutusta, työssä käymistä ja suomen opiskelua. Sen lisäksi tutkimuksessa on vertailtu venäjänkielisten testituloksia muiden osallistujien tekstituloksiin.

Tuloksien mukaan työllistyneet maahanmuuttajat ovat pärjänneet YKI-testissä keskimääräisesti paremmin kuin ne, jotka eivät käy töissä. Kuitenkin työn saaminen vaatii hyvää kielitaitoa hakijalta jo etukäteen. On myös muita tekijöitä, jotka vaikuttavat kielitaidon kehittymiseen, kuten yksilölliset ominaisuudet, motivaatio ja kielenkäyttömahdollisuudet.

Yleinen kielitutkinto on virallinen tapa Suomessa osoittaa kielitaito moniin tarkoituksiin. YKI-tutkinto tarvitaan Suomen kansalaisuuden saamiseksi ja työntantajalle kielitaidon osoittamiseksi, esimerkiksi asiantuntijan ja terveydenhuollon tehtävissä (Valvira.fi). Tämän tutkimuksen informanttien yleisin syy suorittaa YKI-tutkinto oli Suomen kansalaisuuden, työn tai opiskelupaikan hakeminen. Suurin osa työssä olevista pyrkii hakemaan kansalaisuutta, kun taas töitä vailla olevat haluavat saada todistuksen kielitaidosta, jotta työllistyisivät tai saisivat opiskelupaikan Suomessa.

Toisessa luvussa esitän ekologista lähestymistapaa toisen kielen oppimiseen, ja pyrin vastamaan, miten sanastoa opitaan, esittelen sanaston tutkimiseen liittyvät termit ja kuvaan miten sanaston osaamista arvioidaan eurooppalaisen viitekehyksen mukaan. Myöhemmin samassa luvussa pyrin vastaamaan siihen, millainen rooli kieliympäristöllä on kielen oppimisen prosessissa. Kolmannessa luvussa kerron lyhyesti Yleisistä kielitutkinnoista, YKI-korpuksesta ja omasta aineistostani. Esittelen myös, mitä on leksikaalinen diversiteetti. Neljännessä luvussa analysoin YKI-korpuksen aineistoa. Viidennessä luvussa käyn läpi saatuja tuloksia.



## 2 Sanaston oppiminen osana toisen kielen oppimista

### 2.1 Ekologinen lähestymistapa kielen oppimiseen

Termi ”kielelle altistuminen” (eng. *language exposure*) tarkoittaa kielenoppijan kontaktia opittavan kielen kanssa. Kontaktit voivat olla muodoltaan kirjallisia tai suullisia, ja oppijan rooli voi olla passiivinen tai aktiivinen ja tilanne epävirallinen tai virallinen. Toiselle kielelle altistumista tapahtuu kaikkialla ja kenen kanssa tahansa niin kauan kuin oppija voi kuulla, käyttää tai nähdä toista kieltä. Kohdekielelle oppija voi altistua puhumalla muiden ihmisten kanssa ja lukemalla kirjoja tai sanomalehtiä kohdekielellä. Kosketus kieleen voi tapahtua myös viestinnän kanavien kautta, esimerkiksi radion, television, mainoksien ja sosiaalisen median ja muun toiminnan kautta, missä kohdekieltä käytetään. (Magno et al. 2009.)

Psykologi James J. Gibson loi termin affordanssi, eli tarjouma. Gibsonin mukaan, ympäristössä olevat tarjoumat tarjoavat, huolehtivat ja toimittavat eläimille sekä hyvää että huonoa. Ympäristö ja eläin täydentävät toisiaan. Hän myös korostaa, että tarjouma ei ole fyysistä tai ilmiömäistä (Gibson 1979: 127, 143.) Affordanssi on ympäristön ominaisuus, joka on relevantti yksilölle, ja voi olla vaikutukseltaan positiivinen tai negatiivinen. Tarjoumat ovat yksilöllisiä ja riippuvat siitä, mitä sillä hetkellä yksilö tarvitsee, haluaa ja mistä hänelle on sillä hetkellä hyötyä. Aktiivinen kielenoppija havaitsee affordanssit ympäristöstä ja poimii vain ne, joista hänen mielestään on hyötyä. (van Lier 2000: 252.)

van Lier korostaa affordanssien tärkeyttä kielen oppimiselle. van Lierin mukaan ekologisen lähestymistavan keskeinen tutkimuskohde on suhteet ihmisten ja hänen ympärillään olevan maailman välillä. Ympäristöstä löytyy fyysisiä, sosiaalisia ja symbolisia affordansseja, jotka tarjoavat oppijalle mahdollisuuden toimia. van Lierin mukaan maailma on täynnä mahdollisuuksia luoda suhteita. Meidän ympärillämme on fyysisiä asioita, kuten pilviä, puita ja kiviä, sosiaalisia ympäristöjä, kuten taloja, teitä ja luokkahuoneita, sosiaalisia yhteisöjä, kuten perheitä, kouluja ja urheilutiimejä, sekä symbolisia ympäristöjä, kuten tarinoita, historiaa ja uskomuksia, jotka tarjoavat käyttömahdollisuuksia aloittaa toiminta. (van Lier 2004, 2010.)

Ekologisesta näkökulmasta oppiminen vaati oppijalta aktiivista toimintaa. Ihminen ei opi vastaanottamalla syötöstä, joka on hänelle valmiiksi toimitettu, vaan hän oppii ympäristöstä itse poimimalla tietoa oman tarpeensa ja tarkoituksensa mukaan. Sen takia pedagogisesta näkökulmasta katsoen opettajan tehtävänä on ottaa huomioon oppijan tarpeet ja toiminta, jotta oppija pystyy

käyttämään ympäristössä olevat resurssit. Opettajan tehtävä on ohjata oppijan havaintokykyä ja toimintaa, jotta oppija hyödyntää affordansseja ja pystyy näin menemään eteenpäin omien tarpeidensa ja tavoitteidensa mukaan. Havaitseminen on keskeinen oppimisen asia. Ekologisen näkökulman mukaan havaitseminen on moniaistista, ja se voi käsittää ympäröivien maisemien, sanojen ja eleiden havaitsemista. (van Lier 2007: 53–54.)

Leo van Lier pitää kielen oppimista asteittaisena ja kumulatiivisena prosessina, tai tuloksena useista prosesseista, jotka toimivat samanaikaisesti ja peräkkäin. Nämä prosessit ovat altistuminen (eng. *exposure*), työstäminen (*engagement*), sisäistäminen (*intake*) ja taitaminen (*proficiency*). Luokkahuoneessa tapahtuva oppiminen on rajallista, jos oppitunnit ovat ainoa kosketus kieleen sen oppimisprosessin aikana. Vaikka opetusta toteutetaan joka päivä, se ei riitä siihen, että oppijan kieli kehittyisi, tai edistyminen on todella hidasta. van Lierin mukaan opiskelijan täytyy pitää opittavaa kieltä mielessään, ei vain oppituntien aikana, vaan myös oppituntien välillä, eli opetuksen ulkopuolella, jos halutaan, että kehitystä tapahtuu. van Lier toteaa, että kielen kehitys ei tapahdu oppituntien kuluessa, vaan sen sijaan oppituntien välillä. Hänen mukaansa oppitunti antaa opiskelijalle tarvittavat raaka-aineet opiskelua varten, ja on opiskelijan oma valinta, milloin, missä ja miten hän opiskelee myöhemmin kieltä. Ainoastaan kotitehtävien tekeminen ei auta kehittymään, vaan tarvitaan kielen ajattelua ja havaintojen tekemistä kieleen liittyvistä ilmiöistä. (van Lier 1996: 42–43.)

Yleinen käsitys altistumisesta on se, että mitä enemmän oppija altistuu kielelle, sen parempi. Kuitenkin Leo van Lierin mukaan altistumisen laatu on tärkeämpää kuin määrä. Ensiksi kielen pitäisi olla käyttökelpoista, eli käyttäjä pystyy hyödyntämään kieliympäristöstä saatuja tietoja omassa käytössään. Jossain tietyssä tilanteessa jopa pieni määrä kielelle altistumista voi olla tarpeeksi, kun taas jossain ihan eri tilanteessa suuri määrä voi olla täysin hyödytöntä. (van Lier 1996: 45.) Toinen asia, joka vaikuttaa altistumisen laatuun, on avun saatavuus kielenoppimiselle. Tärkeitä ovat kielellinen konteksti (esimerkiksi julkiset kyltit ja tervehdykset) ja kanssakäyminen muiden ihmisten kanssa. Kolmas kielelle altistumisen laatuun vaikuttava tekijä on sosiokulttuurinen ympäristö. Vaikka oppija asuu maassa, jossa hän pääsee altistumaan kielelle, se ei tarkoita sitä, että hän oppii sitä. Jos päivittäin ympäristössä esimerkiksi kotona puhutaan omalla äidinkielellä ja muut asiat on mahdollista hoitaa ilman maan virallista kieltä, ei synny tarvetta oppia uutta kieltä. Tästä kuvaava esimerkki on ensimmäisen polven siirtolaiset. Voimavarojen ja motivaation puute johtavat siihen, että uuden kotimaan kansalliskieli jää opiskelematta, vaikka he altistuvat kielelle päivittäin. (van Lier 1996.)

Kielelle altistuminen itsessään ei riitä kielen oppimiselle, vaan oppijan pitäisi olla vastaanottavainen ja utelias altistavaan kieleen. Nimenomaan vastaanottavaisuus ja uteliaisuus ovat ne ominaisuudet,

jotka van Lierin mukaan auttavat kielen oppimisessa. Leo van Lierin mukaan uteliaisuus ei ole vain passiivista avoimuutta (eng. ”*passive state of openness*”) kielelle, vaan hän kutsuu sitä ”*spirit of exploration*”, joka on lapsella ennen kuin hän pääsee kouluun, mutta jonka koulujärjestelmä tuhoaa. Vastaanottavainen ihminen on sellainen, joka kiinnittää huomionsa kielen aineksiin. (van Lier 1996: 48–49.) Kieliainesten havaitseminen on tärkeä tekijä kielen työstämisen prosessissa. Tämän saavuttamiseksi tarvitaan tietoa, mihin kielen ilmiöihin pitäisi keskittyä. (van Lier 1996: 51.) Kuitenkaan opettajan ei pitäisi määritellä mitä, milloin ja miten opiskelijan täytyy oppia, vaan hänen tehtävänsä on tarjota monipuolisia kieliaiaineita ja antaa oppijan itse valita, mitä hän tarvitsee. Työstäminen tapahtuu silloin, kun oppija on vuorovaikutuksessa ympäristön kanssa, eli kielen ympäristö tarjoaa oppimismahdollisuuksia oppijalle. (van Lier 1996: 52–53.)

Sen jälkeen, kun oppija pystyy tehdä havaintoja kielestä, jolle hän on alttiina, ja keskittyy kielenilmiöihin tai affordanssiin, kielen prosessointi vaatii oppijalta kognitiivista, emotionaalista tai fyysistä vaivannäköä. Sosiaalinen vuorovaikutus ja erityisesti keskustelu ovat luonnollisia ympäristöjä kielen prosessoinnille. Neljäs prosessi, joka vaikuttaa kielitaidon karttumiseen, on taitaminen. Oppilaan pitäisi muistaa se, mitä hän on oppinut, eli tiedot on säilytettävä muistissa sillä tavalla, että ne olisivat helposti saatavilla, kun niitä tarvitaan. Tämän saavuttamiseksi tarvitaan säännöllisiä harjoituksia. (van Lier 1996.)

Ekologista lähestymistapaa on sovellettu suomenkielisissä kielenoppimisen tutkimuksissa. Strömmer väitöskirjassaan (2017) tutkii suomen kielen oppimisen mahdollisuuksia siivoustyössä. Hän on soveltanut ekologista lähestymistapaa suomen kielen oppimisen kontekstiin. Väitöskirjassa kiinnitetään huomiota affordansseihin työympäristössä ja työyhteisön tarjoamaan oikea-aikaiseen tukeen. Väitöskirjassaan Virtanen (2017a) tarkastelee kansainvälisten sairaanhoitajaopiskelijoiden ammatillista kielitaitoa ja toimijuutta sosiokognitiivisessa ja ekologisessa viitekehyksessä.

Tässä tutkimuksessani tarkastelen, miten kohdemaan ympäristö vaikuttaa sanaston oppimiseen ekologisesta näkökulmasta. Suomi toisena kielenä opiskelijoiden ympärillä on huomattavasti enemmän tarjoumia kuin vieraana kielenä opiskelevilla. Henkilö altistuu kielelle joka päivä jonkun verran asumalla kohdemaassa, kuten aiemmin mainitsin (ks. Magno et al. 2009). Altistuminen tapahtuu kaupassa käydessä, kylttejä, ohjeita, mainoksia, työ- tai vuokrailmoituksia lukiessa, reittiä tai apua kysyessä tai lomaketta täyttäessä. Tämän tutkimuksen kohderyhmän oppimisympäristön voi jakaa työympäristöön, vapaa-aikaan, kotiin ja opiskeluympäristöön. Ekologisen näkökulman perusteella jokainen ympäristö voi tarjota erilaisia tarjoumia, joita oppija voi poimia omien tarpeidensa mukaan. Tässä tapauksessa työympäristö voi tarjota uutta sanastoa, joka tarvitaan

tehtävän suorittamiseen, kun taas kahvitauko työkavereiden kanssa tarjoaa mahdollisuuden harjoitella informaalista keskustelua.

## 2.2 Sanaston oppimiseen vaikuttavat tekijät

Tässä tutkimuksessa tutkin sanaston laajuutta kirjoitetuissa teksteissä, eli produktiivista sanaston osaamista. Tämän tutkimuksen kannalta päätin ensiksi lyhyesti selittää reseptiivisen ja produktiivisen sanaston osaamisen välistä eroa ja vasta sen jälkeen käsitellä sanaston oppimiseen vaikuttavia tekijöitä. On hyvä tietää, että kielen oppija ei osaa kaikkia sanoja samalla tavalla, vaan jotkut sanat hän pystyy tunnistamaan tekstistä, kun taas jotkut sanat hän pystyy tuottamaan puheessa tai kirjoituksessa. Reseptiivistä sanavarastoa pidetään produktiivista suurempana. (Pietilä & Lintunen 2014: 143.)

Reseptiivistä sanastoa kutsutaan myös passiiviseksi, kun taas produktiivista kutsutaan aktiiviseksi. Reseptiivinen sanasto viittaa kykyyn muistaa sanan merkitys, kun sanan muoto esiintyy, ja produktiivinen kykyyn tuottaa oikea muoto tarvittavan merkityksen ilmaisemiseksi. Reseptiivinen tieto tarkoittaa kielenkäyttäjän sanojen osaamista lukemisen ja kuuntelemisen aikana, kun taas produktiivinen ilmaisee sanojen käyttöä puheessa tai kirjoituksessa. Reseptiivinen osaaminen saavutetaan ennen produktiivista, koska produktiivinen osaaminen vaati oppijalta enemmän tietoa sanasta, esimerkiksi sen monista merkityksistä, käyttörajoituksista ja kieliopillisista seikoista. (González-Fernández & Schmitt 2017: 284–285.) Produktiivinen sanan osaaminen liittyy sanan syvyyden oppimiseen. Sanaston osaaminen ei tarkoita vain sitä, että kielenkäyttäjä tietää sanan merkityksen ja sen yhteyden muihin sanoihin, vaan hänen täytyy osata tarvittaessa tuottaa sana nopeasti sekä reseptiivisellä että produktiivisella tasolla. (Ellis 2008: 86–87.)

Ellis korostaa, että sanaston ja kieliopin oppimista ei voida erottaa, koska uusien sanojen oppimisen kautta oppija voi oppia kieliopillisia seikkoja. Esimerkiksi oppija voi jakaa uuden muutaman sanan pituisen ilmaisun erillisiin sanoihin ja sen jälkeen tutkia, miten kyseiset sanat erikseen kieliopillisesti käyttäytyvät muiden sanojen tai lauseen yhteydessä. Ellis korostaa, että sanaston oppiminen on enemmän kuin vain uusien sanojen muotojen ja merkityksien oppimista ja niiden painamista muistiin aiemmin opittujen joukkoon. Se tarkoittaa myös syvempää jo opittujen sanojen merkitysten ja käytön oppimista, eli sanan syvyyden oppimista. Tähän sanaston oppimisen prosessiin liittyy myös kieliopin oppiminen. (Ellis 2008: 86–87.)

Ellisin (2008) mukaan toisen kielen oppimiseen voi vaikuttaa monta eri tekijää, kuten psykologiset tekijät, esimerkiksi oppijan kielellinen kyvykkyys, oppimisstrategiat, asenteet, motivaatio, ahdistus,

luonne sekä myös äidinkieli ja ikä. Llanesin ja Muñozin (2013) mukaan myös oppimiskontekstilla on suuri rooli kohdekielen oppimisessa. Seuraavaksi lyhyesti esitän neljä tämän tutkimuksen kannalta relevanttia tekijää: oppijan äidinkieli, ikä, koulutus ja sosiaalinen vuorovaikutus. Myöhemmin alaluvussa 2.3 kerron, miten oppimisympäristö liittyy oppimiseen ja millainen vaikutus toisen kielen oppimiseen sillä on.

### **Äidinkieli**

Termiä *transfer* käytetään, kun puhutaan oppijan äidinkielen siirtovaikutuksesta, eli kielellisten piirteiden siirtymisestä oppijan kohdekieleen. Siirtovaikutus voi olla positiivinen tai negatiivinen. Vanhemmat oppijat voivat hyötyä siirtovaikutuksesta varsinkin, jos äidinkielellä ja kohdekielellä on sukulaissanoja ja jos oppijalla on laaja äidinkielen sanasto. Positiivinen vaikutus on huomattu myös kieliopillisissa seikoissa. Kuitenkin kohdekielen kieliopillisten muotojen erilaisuus voi vaikuttaa negatiivisesti heidän oppimiseensa. Nuorten oppimisessa siirtovaikutus sitä vastoin on epätodennäköisempää. Sitä selittää se, kuinka paljon oppijat ovat riippuvaisia äidinkielestään tai kohdekielestään. Ellis selittää, että kaikilla oppijoilla on kaksi ensisijaista tietolähdettä, kuten äidinkieli ja toinen kieli, jolle he altistuvat. Jos ensimmäisellä kielellä ei ole vielä vahvaa osaamista, oppija luottaa enemmän toiseen kieleen. Se selittää, miksi alle 3-vuotiaat lapset pystyvät oppimaan ääntämään vieraalla kielellä kuten natiivit. (Ellis 2008: 134.) Äidinkielestä voi olla hyötyä ja haittaa uuden kielen sanaston oppimisessa. Esimerkiksi, jos kohdekielen sana muistuttaa oppijan äidinkielessä olevaa sanaa, oppija voi päätellä sen merkityksen äidinkielensä tietojen perusteella. Tällainen tapa voi olla toimiva, mutta joskus samalta suvulta näyttävä sana voi johtaa oppijan harhaan. (Ellis 2008: 86–87.)

### **Ikä**

Vieraan kielen oppimisprosessin alussa vanhemmat oppijat oppivat kieltä nopeammin kuin lapset. Kuitenkin vanhemmat oppijat oppivat nopeammin sanastoa ja kielioppia, mutta eivät ääntämistä. Lapset ajan myötä pystyvät saavuttamaan saman tason kuin aikuisetkin ja jopa ylittämään aikuiset. Näiden kahden ryhmän eroa voi selittää se, että vanhemmat oppijat käyttävät tietoisia oppimisstrategioita, kun taas lasten oppiminen perustuu implisiittiseen oppimiseen. Implisiittinen oppiminen vie aikaa ja vaatii valtavaa määrää altistumista toiselle kielelle, joten edistyminen ei näy niin nopeasti lapsilla kuin aikuisilla. (Ellis 2008: 32–36.)

### **Koulutus**

Sanaston tutkimuksessa koulutusta käytetään muuttujana, koska koulutustasolla on yhteys laajempaan sanastoon. Keuleers et al. (2015) osoittivat tutkimuksessaan, että kasvu sanaston

laajuudessa tapahtuu joka kerta siirryttäessä alemmalta koulutusasteelta korkeammalle, esimerkiksi keskiasteelta kandidaatin tutkintoon ja kandidaatin tutkinnosta maisterin tutkintoon.

### **Sosiaalinen vuorovaikutus**

Suomessa on tehty tutkimuksia, miten suomen kieltä voi oppia työssä tai työharjoittelussa. Sunin (2010) tekemässä tutkimuksessa, joka on osa Suomi työkielenä -hanketta, haastateltiin kahta venäjää äidinkielenään puhuvaa henkilöä. Ensimmäinen heistä on seitsemän vuotta tanssinopettajana Suomessa työskennellyt ja toinen neljä vuotta lääkärinä työskennellyt henkilö. Molemmat ovat muuttaneet Suomeen työn perässä. Tanssinopettajana työskentelevä henkilö käyttää suomen kielen oppimisen apuna omia oppilaitaan, jotka auttavat häntä hänen unohtaessaan oikean sanan. Tutkijan haastattelemat oppilaat olivat yllättyneitä siitä, miten nopeasti opettaja oppi uusia sanoja. Termi *kielen imaiseminen* kuvaa kielen oppimisen prosessin nopeutta ja intensiivisyyttä. Tanssinopettajan kuullessa uuden sanan, hän painaa sen muistiinsa, jotta tulevaisuudessa hän voisi käyttää sitä uudestaan. Suomessa töissä olevien maahanmuuttajien kokemusten mukaan kollegoilta saatu tuki viestintätilanteissa ja kielenoppimisessa on korvaamatonta, ja sitä usein tarvitaan kirjoittamistilanteissa. (Suni 2010: 49, 52–53.) Venäläislääkäri korosti, että potilailta on mahdollista oppia uusia sanoja. Esimerkiksi potilaan puheessa voi esiintyä jokin uusi sana, jonka merkityksen lääkäri selvittää myöhemmin. Tämä tarkoittaa, että lääkäri saa kielen oppimisen mahdollisuuksia töissä. Kysymyksen tai epäselvän kohdan toisto luo mahdollisuuden ymmärtää, ymmärsikö kielenoppija asian oikein, ja myös tehdä toisten sanoista omia. (Suni 2010: 53–54.) Tutkimuksen tulokset osoittivat, että sosiaalisella vuorovaikutuksella on merkitys kielen oppimiselle. Oppija oppi konkreettisia ilmauksia ympärillä olevilta työkavereilta ja muilta ihmisiltä ja informaali sosiaalinen apu auttaa oppimisen etenemisessä. (Suni 2010: 55.)

Samankaltaisia tuloksia on saanut Partanen (2013) omassa tutkimuksessaan, jossa hän on tutkinut millaista työyhteisön tukea ja millaisia mahdollisuuksia suomen kielen oppimiselle voi saada puhdistuspalvelualalla työskentelevät afrikkalaiset maahanmuuttajat. Tulokset osoittivat, että vuorovaikutus työtoverien kanssa ja heiltä saatu kielellinen tuki, kuten toisto tai tekstin sisällön kertominen omin sanoin, ovat auttaneet maahanmuuttajaa kehittämään suomen kielen taitoaan. Tuen saaminen kielenoppimisen prosessissa edistää kielen kehittymistä: ”<...> kielitaito kehittyy nimenomaan toisten tuella, merkityksellisessä ympäristössä toisten sanoja kierrättämällä, niitä jakamalla ja vähitellen omaksi tekemällä” (Virtanen 2017b: 3). Virtasen tutkimus (2017b) osoitti, että sairaanhoitajaksi opiskelevat opiskelijat ovat oppineet suomen kieltä vanhuksilta, eli sairaalaan asiakkailta, koska he ovat toistaneet, selittäneet ja näyttäneet kielellisiä ilmaisuja, kun sairaanhoitajan kielitaito ei ole riittänyt.

### 2.2.1 Sanaston tutkimiseen liittyvä terminologia

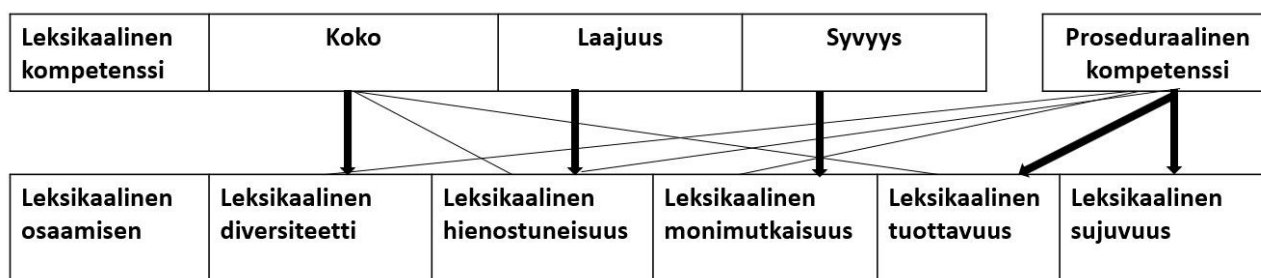
Tässä alaluvussa kirjoitan keskeisistä sanaston tutkimukseen liittyvistä käsiteistä. *Leksikaalinen diversiteetti* (eng. *lexical diversity*) on leksikaalisen tiedon ja taidon indikaattori, ja se kuvaa yksilön sanavaraston laajuutta ja kompleksisuutta (Honko 2017). Sitä pidetään sanojen toistamisen vastakohtana (Jarvis 2013a: 88; 2013b: 15). Sanojen (eng. *type*) ja saneiden (eng. *token*) määrän epätasapaino lauseessa johtuu toistosta. Kun tekstissä on paljon toistoa, se tarkoittaa sitä, että tekstin leksikaalinen diversiteetti pienenee. Kun taas, jos tekstissä on vähän toistuvia sanoja, tekstin leksikaalinen diversiteetti kasvaa. Termillä leksikaalinen diversiteetti ei ole vain yhtä vakituista määritelmää, vaan määritelmä riippuu tutkimuksen tarkastelunäkökulmasta. Kun puhutaan leksikaalisesta diversiteetistä, synonyymina käytetään englanninkielisiä termejä, kuten *lexical variation* (sanojen vaihtelu), *lexical variety* (moninaisuus), *lexical variability* (vaihtelevuus) ja *lexical flexibility* (joustavuus). (Jarvis 2013a: 88; 2013b: 15.) Termin *leksikaalisen diversiteetti* sijaan suomenkielisissä lähteissä käytetään termejä sanaston monimuotoisuus, rikkaus, monipuolisuus ja vaihtelevuus (Honko 2017: 164).

Leksikaalisen diversiteetin ja rikkauden (eng. *lexical richness*) käsitteitä käytetään usein synonyymeinä. Termillä leksikaalinen rikkaus kuvataan sanaston laajuutta, syvyyttä, leksikaalista hienostuneisuutta, tiheyttä ja kaikkea niiden välillä. (Jarvis 2013b: 15, 38.) Jarvisin (2013b) mukaan sanaston rikkaus on termi, johon kuuluu sanaston laajuus ja sen syvyys. Laajuus koostuu sanojen määrästä henkilön aktiivisessa sanastossa. Syvyys tarkoittaa sitä, kuinka hyvin henkilö tuntee tietyn sanan. Oppijan kyky yhdistää sanan muoto sen merkityksen kanssa viittaa laajuuteen. Jotta oppija voisi käyttää sanoja oikein, sujuvasti ja tarkoituksenmukaisesti todellisissa tilanteissa, hänen tietonsa sanoista pitäisi olla monipuolisia. (González-Fernández & Schmitt 2017: 283–284.) Muut tutkijat eivät pidä leksikaalista diversiteettiä leksikaalisen rikkauden alaluokkana, eikä näitä kahta käsitettä tulisi rinnastaa. Tämän takia he ehdottavat, että kielitutkimuksen alalla tutkijat käyttäisivät määritelmää, joka olisi lähellä alkuperäistä määritelmää, kuten Yulen vuonna 1944 tekemä määritelmä, joka on yhteensopiva muiden alojen kanssa. Leksikaalisella rikkaudella Yule tarkoitti sanojen määrää yksilön mentaalileksikossa. (Honko, Jarvis & Vainio 2019: 48; Jarvis 2013b: 15.) Esimerkkinä Jarvis ehdottaa ekologista lähestymistapaa, jossa rikkaus tarkoittaa lajien lukumäärää. Jarvis korostaa, että leksikaalisella diversiteetillä on jo selkeä määritelmä, kun taas termiä leksikaalinen rikkaus ei ole toistaiseksi selvästi määritelty. Hän korostaa, että vaikka leksikaalisella diversiteetillä ja rikkaudella on lähtökohdat, ne kuitenkin pitäisi erottaa. (Jarvis 2013b: 38.)

Vaihteleva terminologia eri lähteissä ja yhden vakiintuneen määritelmän puute tekee sen ymmärtämisestä hankalaa. Tämän takia valitsin Bulté et al. (2008) ehdottaman leksikaalisen

taidokkuuden kehityksen tutkimiseksi tarkoitettua mallia (ks. Taulukko 1). Bulté et al. (2008) mukaan leksikaalinen kompetenssi koostuu kahdesta osatekijästä: deklaratiivisesta ja proseduraalisesta. Deklaratiiviseen kuuluu leksikaalista tietoa (eng. *lexical knowledge*), jonka voi jakaa sanavaraston leksikaaliseen kokoon (eng. *size*), laajuuteen (eng. *width*) ja syvyyteen (eng. *depth*). Koolla viitataan leksikaalisien merkintöjen lukumäärään muistissa, jossa sanojen hallinta vaihtelee sanan tuottamisen ja vastaanoton välillä. Proseduraalinen osatekijä koostuu oppijan taitojen ja tiedon hallinnasta ja se kuvaa, kuinka hyvin tieto kielestä säilyy oppijan leksikaalisessa muistissa. (Bulté et al. 2008: 278–279.) Leksikaalisen kompetenssin sijaan Bulté et al. (2008) ehdottaa käyttämään termiä leksikaalinen taidokkuus (eng. *lexical proficiency*), joka voidaan jakaa viiteen ulottuvuuteen: leksikaaliseen diversiteettiin, hienostuneisuuteen, monimutkaisuuteen, tuottavuuteen ja sujuvuuteen. Leksikaalinen diversiteetti koostuu eri sanojen lukumäärästä, joita oppija käyttää.

Taulukko 1. Bulté et al. (2008) käyttämä analyttinen kehys toisen kielen leksikaalisen taidokkuuden kehityksen tutkimiseksi



Leksikaalinen hienostuneisuus (eng. *sophistication*) tarkoittaa semanttisten suhteiden tuntemusta esimerkiksi synonyymien ja antonyymien suhteen, sekä erilaisten, mutta toisiinsa liittyvien, leksikaalisien vaihtoehtojen käyttöä. Sen takia leksikaalinen hienostuneisuus vastaa sanaston laajuutta. Leksikaalinen monimutkaisuus (eng. *complexity*) viittaa sanaston syvyyteen, ja se vaatii oppijalta osaamista käyttää tietyn sanan harvemmin esiintyviä ominaisuuksia, ei vain prototyyppistä näkökohtaa. Leksikaalinen tuottavuus (eng. *productivity*) tarkoittaa sanamäärää, jonka oppija käyttää tietyn tehtävän suorittamiseen. Leksikaalinen sujuvuus (eng. *lexical fluency*) on kielenkäyttäjän tuottamien sanojen nopeus. (Bulté et al. 2008: 278–280.)

## 2.2.2 Sanaston rooli eurooppalaisessa viitekehyksessä

YKI-tutkinnon kielitaitojen osa-alueen arvioinnissa käytetään 6-portaista taitotasoasteikkoa (ks. Liite 2), joka perustuu eurooppalaiseen viitekehykseen (EVK). Siksi on syytä tarkastella, mikä sanaston asema eurooppalaisessa viitekehyksessä on. EVK on julkaistu vuonna 2001 ja se on tällä hetkellä saatavilla 40 eri kielellä. Vuonna 2018 EVK uudistui. EVK:ta käytetään opetuksen suunnittelun



ohjaamiseen, kuten opetussuunnitelmien, opetusohjelmien, kokeiden ja oppikirjojen laatimisessa. Kielen taitotasojen kuvaukset ovat apuväline opettajille, oppijoille ja työnantajille, kun halutaan saada tietoa oppijan kielen osaamisesta ja kehittymisestä taitotasolta toiselle. Eurooppalaisen viitekehyksen taitotasot ovat seuraavat: peruskielitaito (A1/A2), itsenäisen kielenkäyttäjän kielitaito (B1/B2) ja taitavan kielenkäyttäjän kielitaito (C1/C2). Jokaisella tasolla otetaan huomioon jokainen kielitaidon osa-alue erikseen: kuullun ja luetun ymmärtäminen sekä puhuminen ja kirjoittaminen. (COE, Itsearviointilokerikko.) Suomessa laadittu kuvausasteikko (ks. Kielitaidon tasojen kuvausasteikko) on hieman erilainen kuin alkuperäinen eurooppalainen viitekehys, koska jokainen taso on pilkottu pienempiin alatasoihin, esimerkiksi alin taso A1, jonka osaamista lyhyesti kuvataan suppeana viestintänä kaikkein tutuimmissa tilanteissa, on jaettu kolmeen: A1.1 (Kielitaidon alkeiden hallinta), A1.2 (Kehittyvä alkeiskielitaito) ja A1.3 (Toimiva alkeiskielitaito). Kun taas A2 taso, joka kattaa välittömän sosiaalisen kanssakäymisen perustarpeet ja lyhyen kerronnan, on jaettu kahteen: A2.1 (Peruskielitaidon alkuvaihe) ja A2.2 (Kehittyvä peruskielitaito). B1-tasolla vierasta kieltä osaavan kuvataan selviytyvän arkielämässä, jossa B1.1 on toimiva peruskielitaito ja B1.2 on sujuva peruskielitaito. B2-tason kuvauksen mukaan oppija selviytyy säännöllisessä kanssakäymisessä syntyperäisten kanssa. B2-taso jaetaan kahteen alatasoon: B2.1 (Itsenäisen kielitaidon perustaso) ja B2.2 (Toimiva itsenäinen kielitaito). Tämän kuvausasteikon mukaan korkein taso, jonka oppija voi saavuttaa on C1, jolla hän pärjää monissa vaativissa kielenkäyttötilanteissa ja C1.1 (Taitavan kielitaidon perustaso). Taso C2 ei kuulu kuvausasteikkoon.

Mielenkiintoisia huomautuksia tästä aiheesta on tehnyt Masonen, jonka mukaan eurooppalaisen viitekehyksen itsearviointilokerikossa sana *sanasto* esiintyy vain yhden kerran A2-tasolla: ”Ymmärrän muutamia sanontoja ja kaikkein tavallisinta sanastoa, joka liittyy läheisesti omaan elämäni <...>”, kun taas A1-tasolla sana *sana* esiintyy kaksi kertaa: ”Tunnistan tuttuja sanoja ja kaikkein tavallisimpia sanontoja <...>” ja ”Ymmärrän tuttuja nimiä, sanoja ja hyvin yksinkertaisia lauseita<...>”. Masonen huomaa, että alimmalla tasolla sanaston osaamista pidetään konkreettisena kielitaidon osana, mutta myöhemmin muilla tasoilla ei korosteta sanaston osaamista erikseen, vaan se on ”ikään kuin sulautunut muuhun kielitaitoon eikä siitä – korostetusti – enää erikseen tarvitse välittää”. (Masonen 2013: 188–189.)

Toisaalta aikuisten maahanmuuttajien kotoutumiskoulutuksen opetussuunnitelman perusteiden (2012) kielitaidon tasojen kuvausasteikossa sana *sanasto* ilmenee jopa 19 kertaa kaikissa taitotasoissa. Eniten sanaston osaamista korostetaan suullisen ja kirjoittamistaitojen yhteydessä. Taitotasoilla A1.1 ja A1.2 opiskelija ”osaa hyvin suppean perussanaston”. Taitotasoilla A2.1 ja A2.2 kuvataan sanaston osaamista melko konkreettisesti: ”Osaa helposti ennakoitavan perussanaston <...>” perustarpeisiin liittyvää konkreettista sanastoa <...>”. Tasolla A2.2 korostetaan arkielämään

liittyvän sanaston osaamista: ”*Osa kohtalaisen hyvin tavallisen, jokapäiväisen sanaston <...>*”. Tasolta B1.1 lähtien sanaston osaamista kuvataan abstraktisesti, esimerkiksi ”*Osa käyttää melko laajaa jokapäiväistä sanastoa ja joitakin yleisiä fraaseja ja idiomeja*”. B1.2 ”*Osa käyttää kohtalaisen laajaa sanastoa ja tavallisia idiomeja*.” B2.1 ”*Osa käyttää <...> laajahkoa sanastoa mukaan lukien idiomaattinen ja käsitteellinen sanasto*.” C1.1 tasolla sanasto on sen verran laaja, että ”*rajoittavat ilmaisu erittäin harvoin*.” Aloittamalla hyvin suppeasta sanastosta opiskelijan odotetaan saavuttavaan riittävä sanasto, jotta hän pärjää arkielämässä.

Vuonna 2014 on laadittu ammatilliset suomi toisena kielenä -viitekehykset (Komppa et al. 2014), jotka kattavat tiettyjen korkeakoulutuksen alojen, kuten liiketalouden ja hallinnon (liiketalouden tradenomi), hoitotyön (sairaanhoitaja amk), sosiaalialan (sosionomi amk), tietotekniikan (tietotekniikan tradenomi) ja vanhustyön (sosionomi amk) kielelliset vaatimukset. Kaikki nämä työtehtävät ovat kielellisesti vaativia, koska työt ovat luonteeltaan tiedon tuottamista, käsittelyä ja kommunikointia asiakkaiden kanssa. Viitekehykset kuvaavat, millaista kielitaitoa vaaditaan kussakin näiden alojen työtehtävässä, millaisissa kielenkäyttökonteksteissa, eli missä konkreettisissa paikoissa, sosiokulttuurisessa ja tekstuaalisissa ympäristössä, suomen kieltä käytetään. Sanastolla on tärkeä rooli viitekehyksessä, koska siinä kuvataan konkreettisesti, millaista sanastoa vaaditaan jokaisella alalla. Esimerkiksi terveys- ja hoitoalalla on yksityiskohtaisesti kuvattu, millaista sanaston osaamista sairaanhoitajalta vaaditaan: ”*hallitsee mittayksiköiden käytön/ on tarkka hoitotyölle merkittävissä sanapareissa esim. puutua – puuttua, erittää – eristää/ ymmärtää ja pystyy ilmaisemaan arkaluontoisia asioita kiertoilmauksin*” (Komppa et al. 2014: 28).

## 2.3 Oppimisympäristö ja kielen oppiminen

### 2.3.1 Oppimiskontekstien tyypit ja oppimismuodot

Kielen oppimisympäristöjä voi olla erilaisia. Esimerkiksi suomen kielen oppimiskonteksteja voi olla kahdenlaisia: kohdekulttuurin ulkopuolella tapahtuva opetus, eli kielen opiskelu vieraana kielenä, ja kielen opiskelu kohdekulttuurissa (toisena kielenä). Ulkomailla suomea opiskelevaa voi kutsua ”suomi vieraana kielenä” -opiskelijaksi, kun taas Suomessa suomea opiskelevaa ”suomi toisena kielenä” -opiskelijaksi. Kuitenkin opiskelijan oppimiskonteksti voi muuttua esimerkiksi silloin, kun ulkomailla suomea opiskellut opiskelija tulee Suomeen opiskelemaan. Tämän tutkimukseni kohderyhmä kuuluu suomi toisena kielenä -kontekstiin. Heidän oppimisympäristönsä on Suomi ja sen takia heillä on mahdollisuus oppia kieltä kohdekielisessä ympäristössä, jota pidetään uusien sanojen oppimista nopeuttavana ympäristönä (Pietilä & Lintunen 2014: 150). Kun puhutaan vieraan

kielen oppimisympäristöstä englanninkielisissä lähteissä, käytetään usein Collentine ja Freedin (2004: 155–156) ehdottamaa oppimiskontekstien luokittelua: formaali luokkahuoneopetus (eng. *formal language classroom*), kotimaassa annettu kielikylpyopetus (eng. *domestic immersion*) ja opiskelu ulkomailla (eng. *study abroad*). Kuitenkin Jantunen (2015: 107) toteaa, että tämä luokittelu ei sisällä kohdekulttuurissa tapahtuvaa oppimista, jota hän kutsuu naturalistiseksi oppimiseksi. Hän toteaa, että YKI-testiin osallistuvat myös henkilöt, jotka ovat opiskelleet vähän tai eivät ole koskaan opiskelleet suomea luokkahuoneessa tapahtuvassa opetuksessa, vaan he ”ovat oppineet suomen luonnollisesti”. Luonnollinen oppiminen vaikuttaa sanastoon, koska kielenoppija alkaa käyttää sanoja omasta ympäristöstään. Vertailun vuoksi suomea vieraana kielenä opiskeleva oppii uusia sanoja ainoastaan opettajalta ja oppimateriaaleista. Jantunen huomauttaa, että YKI-testi suoritetaan koetilanteessa, jossa henkilö ei voi käyttää sanakirjaa tai muita apuvälineitä, ja sen takia osallistuja saattaa varmuuden vuoksi käyttää vain niitä ilmauksia, jotka hän osaa hyvin ja on varma, että käyttämällä niitä hän ei tekisi virheitä eli ns. leksikaalisia nallekarhuja (eng. *lexical teddy bears*). (Jantunen 2015: 126–127.) Tutkimukseni nojaa Jantusen edellä ehdottamaan naturalistisen oppimisen ympäristöön. Tutkimukseni kohderyhmällä on kontakti suomen kieleen jollain tavalla. Oppija voi oppia sanoja kahdella tavalla: sattumalta ja tarkoituksellisesti. Ensimmäisellä tavalla oppija oppii kohdekielelle altistumisen kautta, kun taas toisella hän opiskelee sanoja muistamalla. Jotta oppija oppii uuden sanan sattumalta, hänen tarvitsee osata päätellä sanan merkitys kontekstista. Sanat, jotka esiintyvät usein syötöksen yhteydessä, kielen oppija omaksuu nopeammin kuin ne, jotka esiintyvät vain harvoin. (Ellis 2008: 86–87.)

Ellisin mukaan voidaan erottaa kaksi oppimiskontekstia: luonnollinen ja ohjattu. Ensimmäinen tapahtuu jokapäiväisessä ympäristössä, kuten työpaikalla, kotona, netissä, kansainvälisissä konferensseissa ja palavereissa, joissa kohdekieli toimii viestintävälineenä. Toinen tapahtuu esimerkiksi koulussa, yliopistossa ja verkko-opetuksessa. On mahdollista, että osa opiskelijoista kokee kohdekielen oppimista vain luonnollisessa ympäristössä ja toinen osa vain formaalissa opetuksessa. Ellis toteaa, että oppimisvaiheessa moni kokee kohdekielen oppimista molemmissa ympäristöissä. Luonnollisessa ympäristössä tapahtuva oppiminen on epämuodollista, jossa, toisin kuin formaalissa opetuksessa, ei noudateta tarkasti sääntöjä, vaan keskeinen opetuksen periaate on osallistuminen ja havainnointi. Luonnollinen opetus panostaa siihen, mitä on opittu, eikä aiheen hallitsemiseen kuten formaalissa opetuksessa. Luokkahuoneopetusta pidetään opetuksena, joka tapahtuu vain luokkahuoneessa ja jossa opiskelijalla ei ole mahdollisuuksia käyttää kieltä tai mahdollisuuksia on vain vähän luokkahuoneen ulkopuolella. (Ellis 2008: 288, 302.)

Sonya Sahradyanin mukaan formaalin ja informaalin oppimisen välillä on selkeät erot. Formaali opetus on järjestetty ja sisältää oppimistavoitteita, joka tapahtuu formaalissa oppimisympäristössä,

kuten esimerkiksi peruskoulussa. Informaali oppiminen on sitä vastoin spontaanista ja tapahtuu arkipäivän tilanteissa, eikä sisällä oppimistavoitteita. Informaalista oppimista voi tapahtua vapaa-ajalla tai työssä. (Sahradyan 2015.) Non-formaali oppiminen on järjestelmällistä, sillä on rakenne ja jonkinlaiset tavoitteet. Non-formaali koulutus ei johda tutkintoon, mutta voi toimia täydennyskoulutuksen muodossa. Usein museot, kirjastot tai kansalaisopistot voivat tarjota erilaisia suunniteltuja toimintoja, kuten seminaareja, työharjoitteluja ja vapaaehtoistyötä. (Lampinen & Juntunen 2013.)

Tutkimuksessani oppimismuodolla on merkitystä. Osa tutkimukseni informanteista käyttää suomen kieltä lähes päivittäin vapaa-aikanaan, töissä, kotonaan ja opiskelussa. Se tarkoittaa sitä, että heillä on laaja pääsy suomen kielen resursseihin omien tarpeidensa mukaan eri muodoissa. Sahradyanin (2015) tekemä tapaustutkimus osoittaa, että informaalin opetuksen kautta maahanmuuttajat ovat oppineet suomea eniten puhumalla muiden ihmisten esim. ystävien, naapurien, kurssi- ja työkavereiden kanssa. Tutkimukseen osallistuneiden mukaan tällaiset arkipäivän toiminnot, kuten jalkapallon pelaaminen suomalaisten kanssa, television katsominen, radion kuunteleminen sekä kirjojen, lehtien ja mainoksien lukeminen, auttoivat heitä kehittämään suomen kieltä. Erilaiset arkipäivän toiminnot, huolimatta siitä missä ja milloin ne tapahtuvat, ovat vaikuttaneet positiivisesti kuuntelemisen, lukemisen ja puhumisen taitojen kehittymiseen. Kuitenkaan arkipäivän toiminnot eivät vaikuttaneet kirjalliseen taitoon. Aikuiset maahanmuuttajat kehittävät kirjoittamisen taitoa enemmän formaalissa opetuksessa. Kun valitsin tutkittavaksi tätä tutkimusta varten informantit kielenkäytön tilanteen mukaan taustatietolomakkeen perusteella, oletin, että osa informanteista hyödyntää kaikkia näitä informaaleja ja sosiaalisia suhteita harjoitellakseen suomen kieltä arkitilanteissa.

Non-formaalin oppimisen osalta Sahradyanin (2015) tulokset osoittavat, että tutkimukseen osallistuneet ovat kehittäneet suullisia taitojaan seminaareissa ja työpajoilla. Maahanmuuttajat ovat kehittäneet sekä suullisia että kirjallisia taitojaan työharjoittelussa, vapaaehtoistyössä ja myös osallistumalla ammatillisen ja yliopistollisen koulutuksen kursseille, joissa opetuskieli on suomi. Omassa tutkimuksessani työssä oppiminen kuuluu non-formaaliin oppimiseen. Tutkimuksessani oletan, että osa informanteista oppii uutta sanastoa, tai ainakin pääsevät harjoittelemaan suomea tekemällä töitä suomalaisessa työympäristössä, kun taas osalla informanteista ei ole samalaista mahdollisuutta. Vaikka kaikkien informanttien oppimiskonteksti on sama, eli kohdemaassa tapahtuva oppiminen, heidän oppimismuodoissaan on eroja ja tällä on merkitystä tutkimukseni kannalta.

### 2.3.2 Oppimiskontekstin rooli sanaston oppimisessa opiskelu ulkomailla konteksti

Sillä on suuri merkitys, opiskellaanko kieltä luokkahuoneessa kotimaassa, vai ulkomailla kohdemaassa. Kuitenkaan opiskeleminen kohdemaassa itsestään ei auta oppimisessa, jos opiskelija on passiivinen, tai hänellä ei ole sosiaalisia kontakteja, joiden kanssa voisi harjoitella kieltä luokkahuoneen ulkopuolella. Tämän tutkimuksen kannalta on tärkeää ymmärtää, että kielenympäristöllä on merkitys, ja että kielen ympäristö vaikuttaa oppimiseen.

Llanesin and Muñozin (2013) mukaan toisen kielen oppimisprosessiin ja tuloksiin vaikuttavat kaksi keskeistä vaikuttajaa: konteksti ja oppijan ikä. Oppimiskonteksti vaikuttaa oppijan vastaanottaman syötteen laatuun ja määrään oppijaa kohtaan, hänen mahdollisuuksiinsa käyttää kohdekieltä ja opetuksen tapaan. Opiskelu ulkomailla -konteksti on tehokkaampi kuin ”opiskelu kotona” (eng. *study at home*) -konteksti erityisesti suullisten taitojen kannalta. Opiskelu ulkomailla ei vaikuta huomattavasti kirjoitustaitoihin, koska opiskelijoilla ulkomailla on hyvin vähän mahdollisuuksia harjoitella kohdekielen kirjoittamista. Opiskelijat pääsevät harjoittelemaan kieltä enemmän kuuntelemalla ja puhumalla. Opiskelu ulkomailla (eng. *study abroad*) tarkoittaa opetusta, jossa opiskelija kokee molempia, muodollisia ja epämuodollisia, opiskelukonteksteja määrätyn ajan ulkomailla. Oppimista, joka tapahtuu kotimaan koulun tai yliopiston ympäristössä kutsutaan ”opiskeluksi kotona” (eng. *study at home*). (Llanes & Muñoz 2013: 63–64, 79–80.)

On tehty useita tutkimuksia, joissa on verrattu opiskelu kotona ja ulkomailla -konteksteja. Esimerkiksi Collentine (2004) on tutkinut natiiveja englannin puhujia, joista ensimmäinen ryhmä opiskelee espanjaa toisena kielenä muodollisessa ympäristössä Yhdysvalloissa ja toinen ryhmä opiskelee opiskelijavaihdossa Espanjassa. Kun taas Freed, Segalowitz ja Dewey (2004) ovat tutkineet ranskaa opiskelevia kolmessa eri kontekstissa: formaalissa luokkahuoneessa kotimaan yliopistossa, intensiivisellä kesäkurssilla ja ulkomailla.

Foster (2009) on tutkinut opiskelijoita kahdessa eri kontekstissa: eri äidinkieltä puhuvat opiskelijat, jotka asuvat Lontoossa ainakin vuoden verran ja opiskelevat englantia (opiskelu ulkomailla) ja farsia äidinkielenään puhuvat englannin kielen opiskelijat Teheranissa (opiskelu kotona). Vertailun vuoksi hän valitsee kolmanneksi ryhmäksi natiivit Lontoossa asuvat yliopisto-opiskelijat, jotka eivät osaa mitään muuta vierasta kieltä. Tulokset osoittivat, että Teheranissa opiskelevien sanastovalinnat eivät olleet samakaltaisia kuin natiivien, vaan ne olivat rajoitetumpia. Kun taas Lontoossa opiskelevien sanasto oli yhtä monipuolista kuin natiivien. Foster korostaa, että asuminen ja altistuminen kielelle kohdemaassa kaikenlaisissa yhteyksissä päivittäin rikastaa oppijan sanastoa. Asuminen kohdekielen ulkopuolella ja kielen opiskelu vain luokkahuoneessa voi johtaa siihen, että oppija panostaa tuttuihin

sanoihin, käyttää niitä vain rajoitetusti, eikä ota huomioon muita sanojen käyttömerkityksiä. (Foster 2009: 104–105.)

Yliniemi, Dufva ja Mäntylä (2006) ovat tutkineet 3.–5. ja 6.–8.-luokkalaisten käsityksiä vieraan kielen oppimisesta vieraskielisessä ympäristössä. Pietarin suomalaisessa koulussa opiskelevat oppilaat näkivät, kuulivat ja käyttivät venäjän kieltä koulussa ja vapaa-aikanaan. He näkivät venäjänkielisiä tiedotteita ja kylttejä. Oppilaat kuulivat altistavaa kieltä radiosta, televisiosta ja mahdollisesti koulun ulkopuolella. Venäjän kieltä oppilaat pääsivät käyttämään ostoksilla tai heidän tarpeidensa mukaan vapaa-aikana. Kaikki oppilaat, jotka osallistuivat tutkimukseen, olivat sitä mieltä, että kielenoppiminen vieraskielisessä ympäristössä edistyy nopeammin kuin Suomessa, koska ympäristö pakottaa käyttämään vierasta kieltä. Myös keskeiseksi kielenoppimista tukevaksi tekijäksi oli nimetty vuorovaikutus, joka lasten mukaan auttaa kehittämään suullista kielitaitoa.

### 2.3.3 Oppimisympäristö formaalin oppimisen ulkopuolella

Oppimista voi olla kahdenlaista: tietoista eli eksplisiittistä (eng. *conscious learning*) ja tiedostamatonta eli implisiittistä (eng. *learning without attention*). Nämä prosessit voivat myös sekoittua. (Mutta et al. 2017: 184.) Implisiittinen oppiminen tapahtuu ilman tarkoitusta ja ilman tietoisuutta, kun taas eksplisiittinen on tietoista ja tahallista oppimista. Eksplisiittinen oppiminen on pohdintaa kielestä, joka johtaa ymmärtämiseen. (Ellis 2008: 189.) Jantusen mukaan toisen kielen ympäristössä, eli kohdekulttuurissa, tapahtuva kielen oppiminen on implisiittistä ja naturalistista, jota hän myös kutsuu omaksumiseksi. Vieraan kielen oppimisympäristö kuuluu eksplisiittiseen ja luokkahuoneoppimiseen. (Jantunen 2015:108.)

Sylvénin ja Sundqvistin (2017: ii) mukaan koulun ulkopuolella tapahtuva opiskelu voi olla kahdenlaista: opetusohjelmaan kuulumatonta (eng. *extracurricular L2 learning*) ja ulkopuolista (eng. *extramural L2 learning*). Ensimmäinen liittyy yhdellä tai toisella tavalla kouluun ja muodolliseen opetukseen, toinen taas oppijan toimintaan kohdekielen parissa koulun ulkopuolella. Oppijan toiminta ei mitenkään liity opettajan ohjeistuksiin eikä kouluun, vaan oppilas itse vapaaehtoisesti ryhtyy toimintaan.

Kun formaali ja informaali oppimisympäristö sekoittuvat yhteen, syntyy uusia tarjoumia (affordanssi) kielen oppimiselle ja kielenkäytölle. Tällaisessa tapauksessa informaali opetus, joka tapahtuu arkielämässä, lisää oppimiselle formaalissa opetuksessa puuttuvia tarjoumia. Sosiaalisen median käyttö voi luoda mahdollisuuksia käyttää vieraita kieliä koulun ulkopuolella. Näin informaalit oppimisympäristöt teknologian avulla avaavat oppijalle uusia kielenkäytön konteksteja ja luovat

tarjoumia. Mutta et al. (2017) ehdottaa käyttämään termiä *hybridioppiminen* kuvailemaan oppimista, joka tapahtuu sekä formaalissa että informaalissa kontekstissa. (Mutta et al. 2017: 182, 184.)

Runsaasti tutkimuksia on tehty Ruotsissa tietokoneavusteisesta englannin kielen oppimisesta informaalisisessa kontekstissa. Tutkimuksissa tutkittiin, miten informaali englannin kielen oppiminen vaikuttaa kielitaidon kehittymiseen. Esimerkiksi Sylvén ja Sundqvist (2014) ovat tutkineet ruotsalaisten neljäsluokkalaisten (ikä 10–11) englannin oppimista toisena kielenä koulun ulkopuolella ja tietokoneen käyttötottumuksia. Koululaiset altistuvat englannin kielelle sillä tavalla, että koulun jälkeen he katsovat englanninkielisiä elokuvia ja TV-ohjelmia, kuuntelevat musiikkia ja pelaavat pelejä englanniksi. Tutkimus myös osoitti, että koululaiset käyttävät englantia kotonaan huomattavasti enemmän, keskimäärin 7 tuntia per viiko, kuin koulussa. Se on kaksi kertaa enemmän kuin koulussa saatu opetus. Englannin kieltä pelaamisessa käyttävät pojat ovat itsevarmoja heidän puhumisen taidostaan. Myös Sundqvistin (2009, 2011) tutkimustulosten mukaan oppijat, jotka pelaavat tietokonepelejä hallitsevat paremmin englanninkielen sanastoa ja heillä on vahvemmat suulliset taidot verrattuna niihin, jotka eivät pelaa tai pelaavat harvoin. Sylvénin ja Sundqvistin (2012, 2017) tutkimuksien mukaan usein pelaaminen parantaa sanaston lisäksi myös kuuntelun ja lukemisen ymmärtämistä.

Alakouluikäisten suomen oppijoiden suomen kielen käyttö vapaa-ajalla, esimerkiksi kirjojen lukeminen, television katsominen, viestittely, blogien seuraaminen, tietojen etsintä netistä ja suomenkielisten ystävien kanssa ajan viettäminen vaikuttavat positiivisesti sanaston kehittämiseen (Honko 2013: 408). Yläasteella ja lukiossa vieraan kielen ulkopuolisen oppimisen tavat ovat elokuvien ja sarjojen katsominen alkuperäisellä kielellä ilman tekstitystä ja tekstityksellä, ulkomaalaisen musiikin kuunteleminen, kappaletekstien ja kirjojen lukeminen ja matkat ulkomaille. Oppijat itse vapaaehtoisesti luovat affordansseja, jotka edistävät heidän omaa oppimistaan. (Lintunen et al. 2017: 66, 72) Ensimmäiset pari kertaa, kun oppija katsoo elokuvan kohdekielellä, vaikka hän ymmärtää lähes kaikki sanat, hän kuitenkin ei pysy mukana vuoropuhelun nopeuden, ääntämisen ja runsaan syötteen määrän vuoksi. Kuitenkin ymmärtäminen paranee, kun oppija katsoo enemmän ja enemmän elokuvia. Elokuvien katsominen voi olla sopiva tapa kehittää kieltä. (Webb & Rodgers 2009.)

### 3 Aineisto ja menetelmät

#### 3.1 Yleisten kielitutkintojen kuvaus

Yleinen kielitutkinto tunnetaan paremmin lyhenteellä YKI. Yleiset kielitutkinnot ovat aikuisille tarkoitettuja kielitaitotutkintoja. Suomessa Yleisen kielitutkinnon mahdollisia suorituskieliä ovat englanti, espanja, italia, ranska, ruotsi, saame, saksa, suomi ja venäjä. Yleiset kielitutkinnot voi suorittaa kolmella tasolla: perus-, keski- tai ylimmällä tasolla. Arvioinnissa käytetään tasoasteikkoja 1–6. Perustaso (taitotaso 1–2) vastaa eurooppalaisen viitekehyksen tasoja A1 ja A2, keskitaso (taitotaso 3–4) vastaa tasoja B1 ja B2 ja ylin taso (taitotaso 5–6) vastaa tasoja C1 ja C2. Tutkinto koostuu neljästä osakokeesta: puhumisesta, puheen ymmärtämisestä, kirjoittamisesta ja tekstin ymmärtämisestä. Tutkinnon suorittaja saa arvosanan jokaisesta osataidosta. (OPH.) Jos osallistuja saa alle 1 perustasolla, alle 3 keskitasolla tai alle 5 ylimmällä tasolla, tarkoittaa se sitä, että osakoetta ei ole hyväksytty. (Solki.)

Testiin osallistuja itse valitsee, mihin tutkintoon hän osallistuu. Perustutkinnon suorittamista suositellaan niille, jotka käyttävät tutkintokieltä epäsäännöllisesti tai joille perustaidot riittävät. Keskitasoa suositellaan aikuisille, jotka tarvitsevat kieltä jonkin verran sekä vapaa-ajallaan että töissä. Suomen kielen keskitaso ei ole riittävä asiantuntijana toimimiseen. Tutkinnon voi suorittaa kuka tahansa aikuinen valitsemalla hänelle sopivan tason. YKI:n suorittamista varten ei vaadita mitään erityistä opiskeluhistoriaa, eikä tutkinto perustu mihinkään oppikirjaan, joka osallistujan pitäisi lukea. Testissä henkilö näyttää suomen kielen osaamistaan. Testiin osallistuvien henkilöiden suositellaan käydä tutustumassa osakokeiden rakenteeseen, jotta he tietävät millaisia tehtäviä he joutuvat testissä suorittamaan. Osallistuja itse valitsee, mihin kielitutkintoon hän haluaa osallistua. Taso tulee valita 6-portaisen taitotasoasteikon (Liite 2) mukaan ja oman arvion perusteella.

Ensimmäiset YKI-testit seitsemällä kielellä ja kolmella tasolla järjestettiin vuonna 1994. Alussa arvioinnissa käytettiin 9-portaista taitotasoasteikkoa. 9-portainen taitotasoasteikko muutettiin vuonna 2002 6-portaiseksi. (OPH.) Vuoteen 2011 asti tutkinto koostui viidestä osakokeesta: puhumisen, puheen ymmärtämisen, kirjoittamisen, tekstin ymmärtämisen sekä sanaston ja rakenteiden osakokeesta. Vuonna 2011 erillinen sanaston ja rakenteiden osakoe poistettiin ja sen sijaan taidot testataan osana muita osakokeita. (Huhta & Hildén 2016: 18.)

6-portainen taitotasoasteikko, jota käytetään YKI-arvioinnissa, on tehty Euroopan neuvoston yleiseurooppalaisen viitekehyksen taitotasoasteikon mukaan. Sen lisäksi Yleisissä kielitutkinnoissa



käytetään kriteeriviitteistä arviointia, eli testisuorituksia ei verrata keskenään, vaan suorituksia verrataan taitotasokuvausten tasokriteereihin. Kielitutkintojen suorittamisen jälkeen osallistuja saa todistuksen, jossa lukee jokaisen osakokeen taitotasoarvio. YKI järjestetään 72 tutkintopaikassa Suomessa. Tutkintojen kehittämisestä ja testien luomisesta vastaa Opetushallitus sekä Jyväskylän yliopiston Soveltavan kielentutkimuksen keskus (Solki).

### 3.2 YKI-korpus

YKI-tutkinnon suorittamisen yleisin syy on Suomen kansalaisuuden hakeminen. Toinen syy suorittaa YKI on työnhaku tai opiskelu. Kolmas syy voi olla tarve osoittaa kielitaidon taso työnantajalle. Kansalaisuuslain (359/2003) mukaan ”*Kielitaitoedellytyksen täyttäminen voidaan osoittaa suorittamalla yleinen kielitutkinto yleistaitotasolla kolme, valtion kielitutkinto tyydyttävällä suullisella ja kirjallisella taidolla tai perusopetuksen oppimäärä suomi tai ruotsi äidinkielenä.*” Yleisissä kielitutkinnoissa tämä käytännössä tarkoittaa sitä, että vähintään kahdesta osakokeesta pitäisi saada arvosanaksi kolmonen. Kansalaisuuden saaminen vaati sitä, että hakija pystyy kommunikoimaan sekä suullisesti että kirjallisesti, ja siksi Maahanmuuttovirasto hyväksyy vain näiden kolmen osakokeen yhdistelmiä: puhuminen ja kirjoittaminen, puheen ymmärtäminen ja kirjoittaminen tai tekstin ymmärtäminen ja puhuminen. Joka tapauksessa hakijan täytyy osallistua kaikkiin osakokeisiin. Opetushallituksen sivulla lukee, että suomen keskitason kielitutkinnon paikat tulevat täytetyiksi erittäin nopeasti, ja siksi halukkaiden on suositeltavaa ilmoittautua etukäteen. Se osoittaa, että keskitason tutkinto on suosituin. (OPH.)

YKI-korpus on tarkoitettu tutkimuskäyttöön ja se sisältää testiaineistoa kaikista yhdeksästä tutkintokielestä kaikilla tutkintotasoilla. Laaja aineisto korpuksessa on jaettu kahteen: vanhaan ja uuteen. Vanha aineisto on kerätty ennen vuotta 2010 ja uusi vuodesta 2011 lähtien. Uuden ja vanhan keskeinen ero on taustatietolomake, ja siksi aineiston hakeminen molemmista aineistoista samoilla kriteereillä ei ole mahdollista. Molemmissa aineistoissa on vapaaehtoinen taustatietolomake, jonka osallistuja vapaaehtoisesti täyttää tutkinnon suorittamisen yhteydessä. Taustatietolomake (ks. Liite 1) koostuu tiedoista osallistujan äidinkielestä, sukupuolesta, iästä, sosioekonomisesta asemasta, koulutaustasta, ammattialasta, tutkinnon kielen opiskeluajasta ja käyttämisestä sekä tutkintotodistuksen käyttötarkoituksesta. Kvantitatiivisen datan, kuten tasoarvion ja osallistujan taustatietojen, lisäksi korpukseseen kuuluu kolme kirjoittamistehtävää ja yksi tallennettu puhumistehtävä. Kuitenkaan kaikilla osallistujilla ei ole tallennettuna kirjoittamisen tai puhumisen suorituksia, sekä myös taustatietoja puuttuu. Korpuksen aineistojen laajuudet riippuvat sekä

tutkintokielestä että tutkintoon osallistuvien henkilöiden määrästä. Korpukseen lisätään jatkuvasti uutta testiaineistoa.

YKI-korpuksen aineisto on kerätty ja tarkoitettu vain tutkimuskäyttöön. Informanttien luvat tietojen käyttämiseen tutkimustarkoitukseen on kerätty jo ennen suoritusten tallentamista. Korpuksessa testinsuorittajien suoritukset on koodattu id-numeroilla ja tekstinäytteistä ei selviä kirjoittajien henkilöllisyyksiä. Tätä tutkimusta varten on pyydetty aineiston käyttöä varten lupa Soveltavan kielentutkimuksen keskukselta.

### 3.3 Oma aineisto

Valitsin YKI-korpuksesta vanhan aineiston (vuoteen 2010 asti), koska uudessa aineistossa ei ole tallennettuna venäjänkielisten osallistujien kirjoitussuorituksia. YKI-korpuksen haussa voi rajata hakutuloksia siten, että niissä näkyvät vain ne tapaukset, joilla on tallennettuna kirjoittamisen tai puhumisen suoritus. Tällä hetkellä YKI-korpuksen vanhasta aineistosta löytyy 8 349 keskitason venäjänkielisten osallistujien suorituksia ja yhteensä 22 279 suoritusta, joten noin 37 % osallistujista on venäjänkielisiä.

Hakuehtojen avulla on mahdollista suodattaa aineistoa esimerkiksi tutkinnon kielen, tason, tasoarvion ja taustamuuttujien, kuten osallistujan iän, sukupuolen, äidinkielen, peruskoulutuksen, sosioekonomisen aseman ja toimialan mukaan. Halutessaan tutkija voi muokata hakua sisällyttämällä hakuehtoihin tutkinnon kielen opiskeluajan, opiskelupaikan ja kielen käyttötilanteen (kotona, työssä, vapaa-aikana ja opiskelussa). Haun avulla voi saada tietoa esimerkiksi siitä, mistä osallistuja on saanut tietoa Yleisistä kielitutkinnoista, miten hän valitsi tutkintotason ja mihin hän tulee käyttämään tutkintotodistusta.

Taulukko 2. Kohderyhmien valintakriteerit

Kriteerit	Ryhmä A (töissä)	Ryhmä B (ei töissä)
Taso	Keskitaso	
Äidinkieli	Venäjä	
Sosioekonominen asema	<ul style="list-style-type: none"> <li>• Yrittäjä</li> <li>• Ylempi toimihenkilö</li> <li>• Alempi toimihenkilö</li> <li>• Työntekijä</li> </ul>	<ul style="list-style-type: none"> <li>• Työtön</li> <li>• Työvoimakoulutuksen opiskelija</li> </ul>
Pohjakoulutus	Yliopisto /AMK /Ammatillinen oppilaitos / Peruskoulu / Kansakoulu /Lukio /Muu	
Kielenkäyttötilanne	<b>Koti:</b> Lähes päivittäin <b>Työ:</b> Lähes päivittäin <b>Vapaa-aika:</b> Lähes päivittäin <b>Opiskelu:</b> Lähes päivittäin	<b>Koti:</b> En lainkaan <b>Työ:</b> En lainkaan <b>Vapaa-aika:</b> <ul style="list-style-type: none"> <li>• Lähes päivittäin</li> <li>• En lainkaan</li> <li>• Kerran pari viikossa</li> <li>• Kerran kuussa tai harvemmin</li> </ul> <b>Opiskelu:</b> <ul style="list-style-type: none"> <li>• En lainkaan</li> <li>• Lähes päivittäin</li> <li>• Kerran pari viikossa</li> </ul>
Tekstityyppi	Mielipide	

Omina hakukriteereinäni käytin tutkinnon kieltä, tasoa, äidinkieltä, peruskoulutusta, tutkinnon kielen käyttötilanteita ja kirjoittamisen suorituksia (ks. Taulukko 2). Tutkittava tutkintokieli on suomi ja YKI-tutkinnon taso on keskitaso. Osallistujien äidinkieli on venäjä ja heillä on korkean-, keski- tai perusasteen koulutus. Tutkinnon kielenkäyttötilanteet valitsin riippuen ryhmästä. Esimerkiksi ryhmään A kuuluvat vain ne informantit, jotka käyttävät suomen kieltä lähes päivittäin kotona, töissä, vapaa-aikana ja opiskelussa. Ryhmä B eroaa ryhmä A:sta siten, että informantit eivät käytä suomen kieltä töissä eikä kotona, vaan heidän kielenkäyttötilanteensa ovat vapaa-ajan tai opiskelun ympäristöt. Aineistoon kuuluvat vain ne osallistujat, joiden testiaineisto sisältää tallennettuja kirjoittamisen suorituksia.

Aineistoni koostuu 80 lyhyestä mielipidetekstistä. Tekstien pituudet vaihtelevat 51 sanasta 255 sanaan. Tekstien keskipituus on noin 135 sanaa. Tutkittavaksi valitsin yhden kirjoittamistehtävän kolmesta. Tutkimuksessani tutkittavien tekstien tekstilaji on mielipideteksti, ja tarkempi tehtävänotsake on mielipide, joka tehtävätyyppinä on ohjattu tai otsikkopohjainen. Mielipidetekstissä käsitellään jotain ajankohtaista aihetta ja sen tavoitteena on vaikuttaa lukijan käsityksiin tästä asiasta. Kirjoittaja perustelee tekstissä oman mielipiteensä perusteellisesti ja tarvittaessa ilmaisee kritiikkiä,

mutta tekstin tyylin pitäisi olla asiallinen. (Jyväskylän koulutuskuntayhtymä.) YKI-testin tasolla 3 mielipidetekstille on asetettu kriteerit: ”Pystyy tekemään yhteenvetoja sekä esittämään jonkinlaisella varmuudella mielipiteitä kerätystä asiatiedosta”. Tasolla 4: ”Pystyy ilmaisemaan mielipiteensä suhteellisen laajasti sekä korostamaan esityksessään keskeisiä seikkoja.” (Kirjoittamisen arviointikriteerit.)

Jokainen YKI-tutkinnon suorittaja kirjoittaa kolme erityyppistä tekstiä. Valitsin mielipidetekstin, koska se on pisin verrattuna kahteen muuhun tekstiin. Mielipidetekstien aiheet vaihtelevat ja kaikki suorittajat eivät ole kirjoittaneet samasta aiheesta. Korpus ei sisällä tehtävänantoja, ja siksi olen pyrkinyt itse päättämään, millaisesta aiheesta osallistujat ovat joutuneet kirjoittamaan mielipidetekstejä. Teksteistä voi päätellä, että aiheina ovat olleet sellaiset aiheet, kuten miksi ihmisillä on nykyisin niin kiire, mikä hoitopaikka lapsille on paras – oma koti vai päiväkot, onko kauppakeskuksia rakennettu jo tarpeeksi, ovatko koulujen kesälomat väärään aikaan, kumpi on parempi, omistusasunto vai vuokra-asunto ja onko aseiden ostaminen liian helppoa. Ainakin näillä teemoilla molempien ryhmien testin suorittajat ovat kirjoittaneet mielipidetekstejä. Ryhmän A mielipidetekstien eri aiheiden määrä oli suurempi kuin ryhmän B, mutta toisaalta ryhmässä A on enemmän informantteja. Tässä vaiheessa on hyvä muistaa, että YKI-korpuksen aineistot eivät ole samankokoisia. Sen takia ryhmien välinen vertailu mielipidetekstin aiheen mukaan olisi mahdotonta.

### 3.4 Leksikaalinen diversiteetti

Leksikaalista diversiteettiä pidetään positiivisena ominaisuutena, joka on toimivan vuorovaikutuksen keskeinen osatekijä. Kielenkäyttäjän pitäisi osata laajasti sanastoa ja tietää, miten sanoja käytetään, jotta hän pystyy ilmaisemaan tarkan merkityssisällön. Jos käyttäjä hallitsee vain suppean sanaston, hän usein toistaa samoja sanoja omassa puheessaan tai kirjoituksessaan. Siinä tapauksessa jatkuva toisto voi estää toimivan vuorovaikutuksen. Sen lisäksi leksikaalisella diversiteetillä on yhteys tekstin laadukkuuteen, ja leksikaalisen diversiteetin takia teksti voi olla vastaanottajalle kiinnostavampi. (Honko, Jarvis & Vainio 2019: 44.) Leksikaalista diversiteettiä ei käytetä vain kielen oppimisen tutkimuksissa osoittamaan kirjoittamisen laatua tai sanaston osaamista, vaan myös muilla aloilla. On havaittu, että se on indeksi, joka heijastaa puhujakompetenssia, Alzheimerin taudin puhkeamista, kuulon vaihtelua ja sosioekonomista asemaa. (McCarthy & Jarvis 2010: 381.)

Leksikaalista diversiteettiä voi tutkia kahdesta näkökulmasta: tekstin ja yksilön näkökulmasta. Ensimmäisestä näkökulmasta tekstiä voi tutkia yleisellä tai konkreettisella tavalla, kuten rajaamalla tekstin kriteereihin esimerkiksi tekstilajin, tekstin iän tai kohderyhmän mukaan. Yksilön

näkökulmasta voidaan tutkia, miten eri kielen käyttäjät käyttävät kieltä ja verrata heitä keskenään. Leksikaalinen diversiteetti antaa kuvan yksilön sanavaraston laajuudesta, kompleksisuudesta sekä tietoa siitä, kuinka kielellisesti kypsä, luova ja kykenevä kielen käyttäjä on. Se myös mahdollistaa havaintojen tekemiseen kielen käyttäjän kielitaidossa tapahtuvista muutoksista. (Honko, Jarvis & Vainio 2019: 47–48.)

Honko, Jarvis ja Vainio (2019) luokittelevat kuusi keskeistä leksikaalisen diversiteetin osatekijää:

- 1) määrä (eng. *volume*) – saneiden määrä tekstissä;
- 2) vaihtelevuus (eng. *variability*) – sanojen määrä ja sijoittuminen tekstiin;
- 3) jakauman tasaisuus (eng. *eveness*) – sanojen frekvenssijakauma;
- 4) runsaus (eng. *abundance*) – eri sanojen määrä;
- 5) sironta (eng. *dispersion*) – sanojen sijoittuminen tekstiin;
- 6) erityisyys (eng. *disparity*) – sanojen merkitys, funktio ja yksilöllisyys.

Tekstin leksikaalinen vaihtelevuus muuttuu, kun teksti pitenee. Saneiden määrä tekstissä vaikuttaa kaikkiin vaihtelevuuden mittoihin. Vaihtelevuus laskee pidemmässä tekstissä. (Jarvis 2013b: 15, 22.)

Vaihtelevuus mitataan sillä tavalla, että lasketaan kuinka monta sanetta (eng. *tokens*) ja kuinka monta erilaista sanaa (eng. *types*) tekstissä on ja sen jälkeen sovelletaan nämä arvot matemaattiseen yhtälöön. Tämän voi tehdä käsin, mutta teknologian kehityksen myötä mittaus nykyisin tehdään automattisesti tietokoneella varsinkin silloin, kun pitäisi käsitellä isoja aineistoja. Sitä varten on kehitetty leksikaalisen diversiteetin mittareja, kuten vocd-D, MTL D ja HD-D. (Jarvis 2017: 538.) MTL D-indeksi on ainoa, johon ei vaikuta otoskoko eikä jakauman tasaisuus (Jarvis 2013b: 28).

Suomenkielisiä tutkimuksia leksikaalisesta diversiteetistä on hyvin vähän. Pääsääntöisesti on tutkittu kirjoitettua kieltä sekä kertovaa tekstiä. Malin (2012) käytti pro gradu -tutkielmassaan TTR-sanatoisteisuuslukua, MTL D-rikkauslukua ja Shannonin indeksiä. Hänen aineistonsa koostuu aikuisten Yleisten kielitutkintojen ja Cefling-hankeen yläkoululaisten kirjallisista suorituksista. Tutkimuksessaan Malin vertaa aikuisten ja lasten sanastoa taitotasolta toiselle siirryttäessä sekä muodollisten ja epämuodollisten tekstilajien vaikutusta sanaston rikkauteen. Laine-Leinosen (2013) pro gradu -tutkielmassa käytetään Shannonin indeksiä ja muita rikkauslukuja. Tutkimuksen kohderyhmä on 1.–6.-luokkalaisten ja aineisto on Koulukorpus 2007–2011. Tutkimuksessa hän vertaa tyttöjen ja poikien sanastoa keskenään sekä sanaston kehittymistä vuosiluokkien välillä. Tulosten mukaan hypoteesi, että tyttöjen sanasto on monipuolisempi kuin saman ikäisten poikien sanasto, pitää paikkansa. Honko (2013) tutkii väitöskirjassaan toisen sukupolven maahanmuuttajien ja peruskoulua käyvien S1-verrokkien leksikaalista tietoa ja taitoa. Menetelmänä hän käyttää SOP-diversiteetti-indeksiä (eng. *sums of probabilities*). Myöhemmin Honko (2017) on tutkinut puhutun kielen

leksikaalista diversiteettiä käyttämällä samaa SOP-indeksiä. Hakala (2018) on tutkinut pro gradu -tutkielmassaan ulkomailta Suomeen tulleiden lääkäreiden sanastoa Tampereen yliopiston järjestämässä laillistamiskuulustelussa. Hänen käyttämänsä menetelmät ovat Shannonin indeksi ja MTLD-rikkausluku.

### 3.5 MTLD leksikaalisen diversiteetin mittarina

Tässä tutkimuksessa käytän mittarina MTLD:tä (eng. *Measure of Textual Lexical Diversity*). MTLD-mittaria pidetään indeksinä, jonka tuloksiin ei vaikuta tekstin otoskoko (Jarvis 2013a: 94). MTLD-mittari toimii niin, että se laskee keskimääräisen peräkkäisten saneitten määrän tekstissä, jotka ylläpitävät annetun TTR-arvon (eng. *type/token ratio*), joka tässä tapauksessa on 0.72. McCarthy ja Jarvis (2010) käyttävät selittääkseen, miten MTLD-rikkausluku toimii, suosittua Lincolnin ilmausta: ”*of the people, by the people, for the people*”. Tässä esimerkissä jokainen tekstin sana arvioidaan peräkkäin sen TTR-arvon saamiseksi: *of (1.00) the (1.00) people (1.00) by (1.00) the (0.800) people (0.667) for (0.714) the (0.625) people (0.556)* jne. Jokainen uusi sana saa arvon 1. Sen jälkeen toistuvat sanat saavat pienemmän arvon kuin yksi niin kauan, kunnes TTR-arvo laskee alle 0.72. Tässä tapauksessa se on sanan *people (0.667)* kohdalla. Silloin kun annettu arvo on saavutettu, kokonaisfaktori kasvaa yhdellä (1) ja TTR arviointi nollataan, eli aloitetaan TTR:n arviointi uudestaan seuraavasta sanasta. Tällä tavalla em. esimerkki lasketaan näin:

*of (1.00) the (1.00) people (1.00) by (1.00) the (.800) people (.667) ||| **Faktori = Faktori +1** ||| for (1.00) the (1.00) people (1.00) ja jne.*

Tässä esimerkissä kokonaisfaktorin lukumäärä on yksi, koska TTR-arvo on vain yhden kerran laskenut alle 0.72, jolloin TTR-arviointi on aloitettu uudestaan. Sen jälkeen lasketaan häntäfaktori jäljelle jääneille sanoille, eli usein niille sanoille, jotka eivät muodosta kokonaisia faktoreita ja ovat jääneet viimeiseksi tekstissä. Esimerkiksi jos tekstiin jää yksi sana, jonka TTR-arvo on 0.887, se tarkoittaa sitä, että kokonaisfaktorin (1) ja TTR-arvon (tässä tapauksessa 0.72) välinen etäisyys lasketaan näin:  $1 - 0.887 = 0.113$ . Kun jäljelle jääneen sanan arvo laskettu, pitäisi laskea, kuinka kaukana se on kokonaisfaktorista, jonka arvo lasketaan  $1 - 0.72 = 0.28$ . Kun molemmat arvot on saatu, lasketaan prosentuaalinen osuus, eli kuinka paljon pienempi jäljelle jääneen sanan TTR-arvo on verrattuna kokonaisfaktorin (1). Tämän edellä mainitun sanan prosentuaalinen osuus lasketaan  $0.113 \cdot 100\% \div 0.28 = 40,357$ . Se tarkoittaa sitä, että sanan, jonka TTR-arvo on 0.887, lasketuksi etäisyydeksi kokonaisfaktorista (0.113) muodostuu noin 40,4 % kokonaisfaktorin arvosta (0.28).

Esimerkiksi jos tekstissä on neljä faktoria ja yksi jäljelle jäänyt sana, jonka TTR-arvo on 0.887, silloin lopullinen faktori lasketaan näin:  $4.00$  (kokonaisfaktori)  $+ 0.404$  (häntäfaktori)  $= 4.404$

Kun kaikki nämä arvot on laskettu, seuraava vaihe on laskea, kuinka monta sanaa tekstissä on ja jakaa saatu määrä lopullisella faktorilla. Esimerkiksi jos teksti koostuu 340 sanasta, ja sen lopullinen faktori on aiemmin laskettu 4.404, silloin MTLN-arvo on 77.203. MTLN-arvoja lasketaan kaksi, joista ensimmäinen lasketaan etuperin ja toinen takaperin. Lopullinen arvo on MTLN-arvo. (McCarthy & Jarvis 2010: 384–385; Malin 2012: 28–29.)

### 3.6 Ohjelmat

Tilastolliset laskelmat, kuten lineaarisen regressiomallin rakentaminen, on tehty R-ohjelmointiympäristön versiolla 3.6.1. MTLN-keskiarvot on syötetty Excel-  
taulukkolaskentaohjelmistoon, jonka avulla on tehty keskiarvojen, mediaanien, minimi- ja maksimi-  
arvojen laskeminen ja kaavioiden piirtäminen. YKI-korpuksesta poimittu aineisto, on käsitelty UDPipe -ohjelman avulla (Straka & Straková 2017). Se on koulutettava ohjelma, jonka avulla automaattisesti onnistuu sanojen lemmaaminen, saneistus, perusmuodistaminen ja riippuvuuksien jäsentely. Ohjelman avulla tekstien sanat muutetaan perusmuotoihin, jotta eri taivutusmuodot lasketaan saman sanan esiintymiksi. Se käsittelee tekstejä CoNLL-U formaatilla. (UDPipe.) Ennen aineiston syöttöä sovellukseen käsittelen informanttien tekstit korjaamalla kirjoitusvirheet. Usein nämä virheet liittyvät astevaihteluun, taivutukseen sekä vokaaleihin *a* ja *ä*. Kun tekstit on korjattu, siirrän jokaisen tekstin erilliseen tekstitiedostoon ja sen jälkeen tekstit ovat valmiita ohjelman käsiteltäväksi.

## 4 Aineiston analyysi

### 4.1 Muuttujat

Tässä tutkimuksessa tarkastelen aineistoani seuraavien muuttujien perusteella: peruskoulutus, sosioekonominen asema, ikä, tekstin pituus, kielenkäyttötilanne (ks. Taulukko 3). Muuttujat valitsin YKI-tutkinnon taustatietolomakkeen perusteella, josta saa tietoa myös muista taustamuuttujista, kuten sukupuolesta ja toimialasta. Kaikkien informanttien äidinkieli on sama, eli venäjä. Olen jakanut ryhmät pohjakoulutuksen mukaan: töissä vai ei, korkeakoulutetut ja ei-korkeakoulutetut ja missä kielenkäyttötilanneissa suomen kieltä käytetään. Sen lisäksi olen laskenut informanttien iän ja tekstien pituuksien keskiarvot ja mediaanit. Tässä alaluvussa lyhyesti kerron, miten informantit jakautuvat aineistossani ja ryhmien sisällä. Tässä tutkimuksessa ryhmä A viittaa informantteihin, jotka testin suorituspäivänä olivat työllistyneitä, eli heillä oli työpaikka, ja ryhmä B koostuu informanteista, jotka ovat merkinneet taustatietolomakkeeseen, että he olivat työtä vailla.

*Taulukko 3. Aineisto eri taustamuuttujien mukaan jaoteltuna*

Muuttujat	Ryhmä A		Ryhmä B	
Sosioekonomi- nen asema	Töissä <sup>3</sup>		Työtön <sup>4</sup>	
Peruskoulutus	Korkeakoulutetut <sup>1</sup> 23 kpl	Ei- korkeakoulutetut <sup>2</sup> 28 kpl	Korkeakoulutetut 15 kpl	Ei-korkeakoulutetut 14 kpl
Ikä	mediaani: 35 keskiarvo: 36,65	mediaani: 36 keskiarvo: 35,46	mediaani: 33 keskiarvo: 31,87	mediaani: 33 keskiarvo: 34,29
Tekstin pituus (sanaa)	min: 51 mediaani: 135 keskiarvo: 137,84 maks: 249		min: 62 mediaani: 134 keskiarvo: 131,38 maks: 255	
Kielenkäyttö- tilanne <sup>5</sup> : kotona, töissä, vapaa-aikana, opiskelussa	<b>Koti:</b> Lähes päivittäin <b>Työ:</b> Lähes päivittäin <b>Vapaa-aika:</b> Lähes päivittäin <b>Opiskelu:</b> Lähes päivittäin		<b>Koti:</b> En lainkaan <b>Työ:</b> En lainkaan <b>Vapaa-aika:</b> <ul style="list-style-type: none"><li>• Lähes päivittäin</li><li>• En lainkaan</li><li>• Kerran pari viikossa</li><li>• Kerran kuussa tai harvemmin</li></ul> <b>Opiskelu:</b> <ul style="list-style-type: none"><li>• En lainkaan</li><li>• Lähes päivittäin</li><li>• Kerran pari viikossa</li></ul>	



<sup>1</sup> Korkeakoulutetut – ne, jotka ovat suorittaneet yliopiston tai ammattikorkeakoulun tutkinnon.

<sup>2</sup> Ei-korkeakoulutetut – ne, jotka ovat suorittaneet ammatillisen oppilaitoksen/ Peruskoulun/ kansakoulun/ Lukion/ Muun

<sup>3</sup> Töissä – Yrittäjä, Ylempi toimihenkilö, Alempi toimihenkilö, Työntekijä

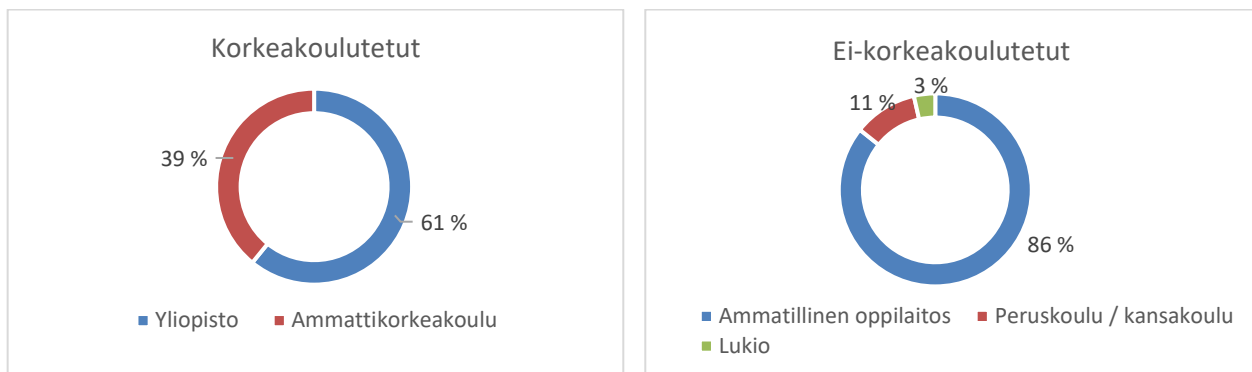
<sup>4</sup> Työtön – ne, jotka testin suorittamispäivänä olivat työttömiä

<sup>5</sup> Kielen käyttö tilanne – missä ja kuinka usein informantit käyttävät testattavaa kieltä

## **Taustamuuttujat**

**Pohjakoulutus.** Yhteensä ryhmässä A on 51 ja ryhmässä B 29 informanttia. Ryhmä A (ks. Kuvio 1) koostuu 23:sta korkeakoulutetusta ja 28:sta ei-korkeakoulutetusta töissä olevasta YKI-tutkinnon suorittajasta. Heistä 14 on suorittanut yliopiston, 9 ammattikorkeakoulun, 24 ammatillisen oppilaitoksen, 1 lukion ja 3 perus-/kansakoulun.

*Kuvio 1. Ryhmän A jakauma koulutuksen mukaan*



Ryhmä B (ks. Kuvio 2) vastaavasti koostuu 13:sta yliopiston, 2:sta ammattikorkeakoulun, 8:sta ammatillisen oppilaitoksen, 4:sta perus-/kansakoulun, 1 lukion ja 1 muun suorittajasta.

*Kuvio 2. Ryhmän B jakauma koulutuksen mukaan*

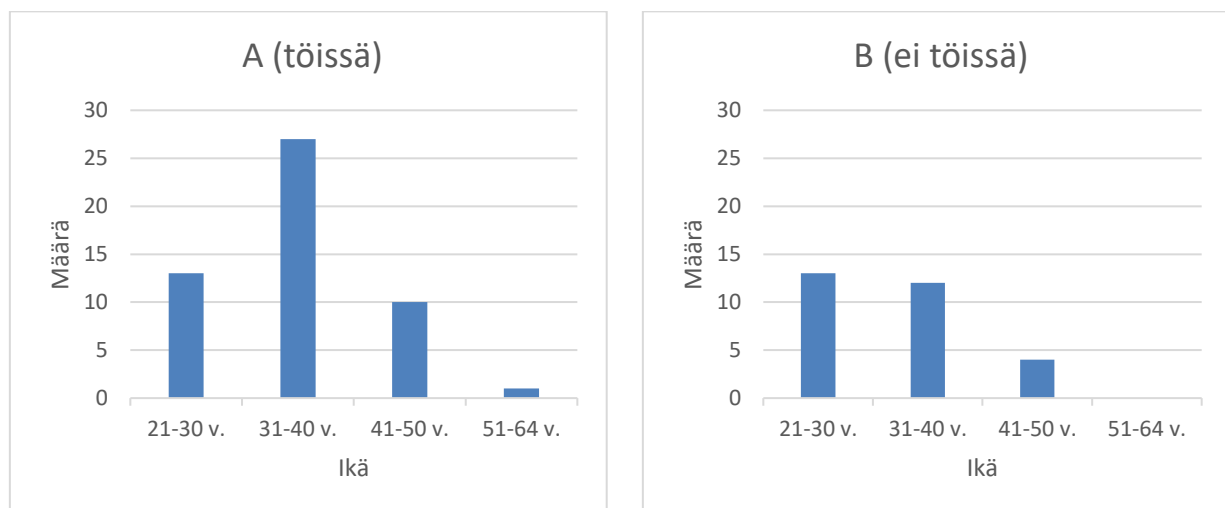


**Sosioekonominen asema.** Ryhmän A informantit YKI-tutkinnon suorituspäivänä taustatietolomakkeen perusteella olivat töissä. Töissä olevaksi lasketaan ne henkilöt, jotka taustatietolomakkeessa merkitsivät omaksi ammatinimikkeekseen jonkin seuraavista: ylempi toimihenkilö, alempi toimihenkilö, työntekijä tai yrittäjä. Tässä tutkimuksessa työttömien ryhmään (B) kuuluvat ne, jotka taustatietolomakkeessa sosioekonomisen aseman kohdalle merkitsivät

olevansa työttömiä tai työvoimakoulutuksen opiskelijoita. Työ- ja elinkeinotoimiston mukaan ammatillinen työvoimakoulutus kuuluu työttömille tai työttömyysuhan alaisille aikuisille, jotka ovat suorittaneet oppivelvollisuutensa (Te-palvelut). Tässä aineistossa töissä olevien määrä on huomattavasti suurempi (51), kuin töitä vailla olevien (29). Prosentuaalisesti töissä käyvien määrä tässä aineistossa on 63,75 % ja työtä vailla olevien 36,25 %.

**Ikä.** Kuten voi nähdä Kuvioista 3, molemmissa ryhmissä on 13 informanttia, jotka ovat 21–30-vuotiaita. Ryhmässä A on yli kaksinkertaisesti enemmän informantteja, jotka ovat 31–40-vuotiaita, kuin ryhmässä B (ryhmässä A 27 kpl ja ryhmässä B 12 kpl). Ryhmässä A on myös kaksi kertaa enemmän 41–50-vuotiaita. Ryhmässä A on yksi yli 51 vuotta vanha informantti, kun taas ryhmässä B ei ole tämän ikäisiä lainkaan. Tässä tutkimuksessa nuorin informantti on 22 vuotta vanha ja vanhin 56-vuotias.

*Kuvio 3. Ryhmän A ja B ikäjakauma*



Ryhmän A korkeakoulutettujen ja töissä olevien iän mediaani on 35 (ei-korkeakoulutettujen 36) ja iän keskiarvo on 36,65 (ei-korkeakoulutettujen 35,46). Ryhmän B sisällä sekä korkeakoulutettujen että ei-korkeakoulutettujen iän mediaani on 33, ja korkeakoulutettujen iän keskiarvo on 31,87 (ei-korkeakoulutettujen 34,29). Ryhmän A korkeakoulutettujen minimi-ikä on 24 ja maksimi-ikä on 56, kun vastaavasti ryhmän B minimi-ikä on 22 ja maksimi-ikä 45. Mielenkiintoista on se, että sekä ryhmän A että B ei-korkeakoulutettujen minimi-ikä on 23 ja maksimi-ikä on 50.

### **Muut muuttujat**

**Tekstin pituus.** Ryhmän A lyhin teksti on 51 sanaa ja pisin 249 sanaa, kun taas ryhmän B minimi tekstin pituus on 62 ja maksimi 255 sanaa. Ryhmän A tekstien pituuksien keskiarvo on noin 138 sanaa ja B ryhmän 131 sanaa. Tekstin pituuksien osalta en huomannut suurempaa eroa, jonka perusteella nämä kaksi ryhmää voisi erottaa.

**Kielenkäyttötilanne.** Taustatietolomakkeen perusteella ryhmän A informantit käyttävät suomen kieltä kotonaan, töissä, vapaa-aikana ja opiskelussa lähes päivittäin, eli he ovat täysin altistuneet suomen kielelle. Ryhmän B informantit ovat työtä vailla olevia, ja sen takia heillä ei ole mahdollisuuksia käyttää kieltä päivittäin töissä, eivätkä he taustatietolomakkeen perusteella käytä kieltä kotonaan, mutta he mahdollisesti käyttävät suomen kieltä opiskelussa ja vapaa-ajalla. Heidän kielenkäyttötilaisuutensa ovat vähäisiä verrattuna ryhmään A, jonka informantit käyttävät suomen kieltä lähes päivittäin kaikissa tilanteissa.

Taulukossa 4 olen jakanut ryhmän B informantit käyttötilanteen mukaan. Taulukosta voimme nähdä, kuinka usein ja missä tilanteessa ryhmän B YKI-tutkinnon osallistujat käyttävät testattavaa kieltä. Eniten on sellaisia informantteja, jotka käyttävät suomen kieltä lähes päivittäin opiskelussa, mutta vain kerran pari viikossa vapaa-aikana. He muodostavat yli puolet (17/29) kaikista ryhmän B informanteista. Kolme informanttia käyttää suomen kieltä opiskelussa, mutta ei vapaa-aikana, kun taas viisi informanttia sitä vastoin ei käytä suomen kieltä opiskelussa, mutta käyttää lähes päivittäin vapaa-aikana. Muut eivät käytä ollenkaan tai harvoin. Keskeinen ero ryhmien välillä on se, että ensimmäisellä ryhmällä on mahdollisuus päivittäin käyttää kieltä sekä työssä että kotona ja toisella ryhmällä tällaista mahdollisuutta ei ole.

Ryhmällä B on rajoitettu määrä tilanteita, joissa he voisivat altistua suomen kielelle. Sen takia halusin tarkastella tarkemmin, miten ryhmän B sisällä kielenkäyttötilanteet informanttien kesken eroavat. Kuitenkaan ryhmän B sisällä ei huomattu suurempaa eroa. Sekä korkeakoulutetut että ei-ke korkeakoulutetut käyttävät kieltä lähes saman verran. Suurin osa käyttää sitä opiskelussa lähes päivittäin. Muiden informanttien pienen lukumäärän takia en voi tehdä johtopäätöksiä ryhmän B sisällä.

*Taulukko 4. Ryhmän B jakautuminen kielenkäyttötilanteen mukaan*

Vapaa-aika	Opiskelu	B kk <sup>1</sup>	B ek <sup>2</sup>
kerran pari viikossa	lähes päivittäin	7	10
en lainkaan	lähes päivittäin	1	2
lähes päivittäin	en lainkaan	4	1
kerran kuussa tai harvemmin	en lainkaan	1	0
kerran pari viikossa	kerran pari viikossa	1	0
en lainkaan	en lainkaan	1	1

<sup>1</sup>kk – Korkeakoulutetut, <sup>2</sup>ek – Ei-ke korkeakoulutetut

Valitettavasti YKI-tutkinnon taustatietolomakkeen perusteella ei saa tietoa, millaisissa konkreettisissa tilanteissa YKI-suorittajat käyttävät suomen kieltä vapaa-aikanaan. Tämän tutkimuksen kannalta olisi ollut mielenkiintoista tietää, käyvätkö he harrastuksissa suomalaisten kanssa, katsovatko he suomalaisia TV-ohjelmia, elokuvia tai sarjoja tai lukevatko he suomalaisia sanomalehtiä tai kirjoja, seuraavatko he suomenkielistä sosiaalista mediaa tai kuuntelevatko he suomalaista musiikkia. Nämä asiat voivat vaikuttaa syötteen määrään ja laatuun. Taustatietolomake, jota on käytetty tässä tutkimuksessa, antaa vain abstraktia tietoa tilanteista, missä ja missä määrin kieltä käytetään. Vapaa-aika on laaja termi ja se voi sisältää harrastuksia, television katsomista, median seuraamista, ajanviettoa ystävien kanssa jne.

## 4.2 Aineiston MTLN-arvojen vertailu

Tässä tutkimuksessa leksikaalinen diversiteetti on laskettu MTLN-mittarin avulla, jota myöhemmin sovelletaan lineaariseen regressiomalliin. Koska tämän tutkimuksen tarkoitus on tutkia leksikaalista diversiteettiä, päätin vertailla keskenään molempia ryhmiä saatujen MTLN-arvojen ja taustamuuttujien perusteella. Tässä vertailussa kiinnitän huomiota MTLN-keskiarvoihin ja -mediaaneihin. Vertaan MTLN-keskiarvoja sosioekonomisen aseman (A ja B) ja koulutuksen (korkeakoulutetut (kk) ja ei-korkeakoulutetut (ek)) perusteella. Suurempi MTLN-arvo osoittaa suurempaa sanastoa.

### **MTLN ja työssä käyminen**

Tässä tutkimuksessa minua kiinnostaa tietää, miten työympäristö vaikuttaa sanaston laajuuteen. Hypoteesini on, että työssä olevien informanttien sanasto on laajempi kuin niiden, jotka eivät käy töissä.

*Taulukko 5. MTLN-keskiarvojen jakautuminen työllistymistilanteen mukaan*

	<b>Ryhmä A</b>	<b>Ryhmä B</b>	<b>Ero</b>	<b>%</b>
<b>MTLN-keskiarvo</b>	<b>60,11</b>	<b>47,53</b>	<b>12,58</b>	<b>20,92</b>
MTLN-mediaani	56,03	44,19	11,84	21,13

A ja B ryhmien MTLN-keskiarvojen vertailusta (ks. Taulukko 5) paljastuu, että työssä käyminen vaikuttaa positiivisesti sanastoon. Se voi johtua työssä tapahtuvasta vuorovaikutuksesta. Työssä käyvillä on mahdollisuus altistua kielelle ja toimia aktiivisena kielenkäyttäjänä. Ryhmän A MTLN-keskiarvo (60,11) on huomattavasti suurempi kuin ryhmän B (47,53), ja sen takia tulos on tämän tutkimuksen kannalta erittäin merkittävä.

## MTLD ja pohjakoulutus

*Taulukko 6. MTLD-keskiarvojen jakautuminen pohjakoulutuksen mukaan*

Ryhmä	KK <sup>1</sup>	EK <sup>2</sup>
<b>A</b>	MTLD-keskiarvo: <b>59,11</b> MTLD-mediaani: 55,27	MTLD-keskiarvo: <b>60,93</b> MTLD-mediaani: 59,39
<b>B</b>	MTLD-keskiarvo: <b>50,13</b> MTLD-mediaani: 44,43	MTLD-keskiarvo: <b>44,74</b> MTLD-mediaani: 42,64
<b>Yhteensä</b>	MTLD-keskiarvo: <b>55,57</b> MTLD-mediaani: 51,35	MTLD-keskiarvo: <b>55,54</b> MTLD-mediaani: 51,53

<sup>1</sup>kk – Korkeakoulutetut, <sup>2</sup>ek – Ei-korkeakoulutetut

Kun vertailin molempien ryhmien informantteja koulutuksen mukaan, eli korkeakoulutetut ja ei-korkeakoulutetut, huolimatta siitä, ovatko he töissä tai eivät, molempien ryhmien MTLD-keskiarvot ovat lähes samat (ks. Taulukko 6). Sen takia voin väittää, että korkeakoulutus yksinään ei vaikuta MTLD-arvoon. Kuitenkin oletan, että korkeakoulutettujen sanasto on laajempi jo alun perin. Sen takia päätin tarkastella, miten MTLD-arvot käyttäytyvät A ja B ryhmien sisällä.

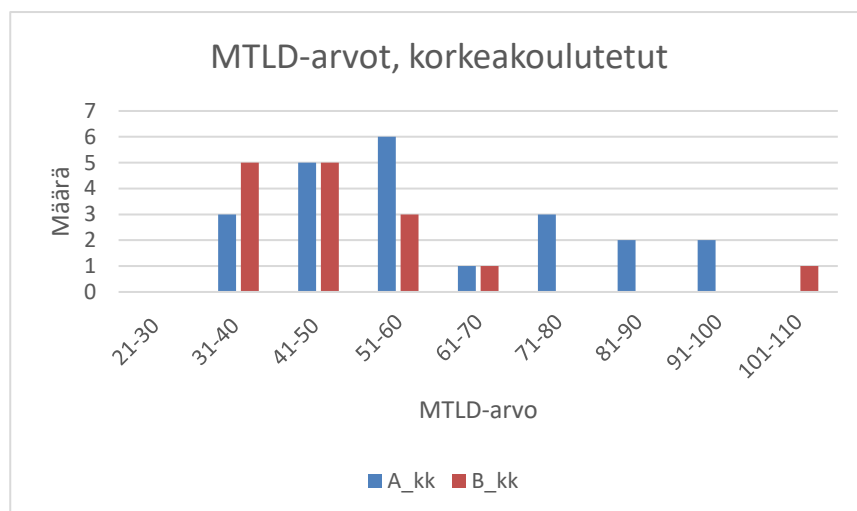
Kun vertailin informanttien MTLD-arvoja ryhmän A sisällä, kävi ilmi, että töissä käyvien ei-korkeakoulutettujen MTLD-keskiarvo on hieman suurempi kuin korkeakoulutettujen (A\_kk 59,11 ja A\_ek 60,93). Tämän eron voisi selittää se, että ei-korkeakoulutetut työssä käyvät oppivat uutta sanastoa, ja sen takia heidän sanastonsa laajenee niin paljon, että se jopa heijastuu MTLD-arvoihin.

Kävi ilmi, että korkeakoulutuksella on merkitystä silloin, kun kielenoppija ei käy töissä eikä hänellä ole korkeakoulutuskintoa. Se heijastuu ryhmän B MTLD-arvoihin, sillä korkeakoulutettujen informanttien MTLD-keskiarvo on huomattavasti korkeampi kuin ei-korkeakoulutettujen (B\_kk 50,13 ja B\_ek 44,74). Tämä osoittaa, että ei-korkeakoulutettujen, työssä käymättömien sanasto on suppeampi kuin korkeakoulutettujen työtä vailla olevien.

Kun vertailin työssä käyviä korkeakoulutettuja ja työtä vailla olevia korkeakoulutettuja, sain vahvistuksen jo alussa saatuun tulokseen, että töissä käyminen nostaa MTLD-arvoa huomattavasti (A\_kk 59,11 ja B\_kk 50,13). Samankaltaiset tulokset sain töissä käyvien ja työtä vailla olevien ei-korkeakoulutettujen vertailussa (A\_ek 60,93 ja B\_ek 44,74).

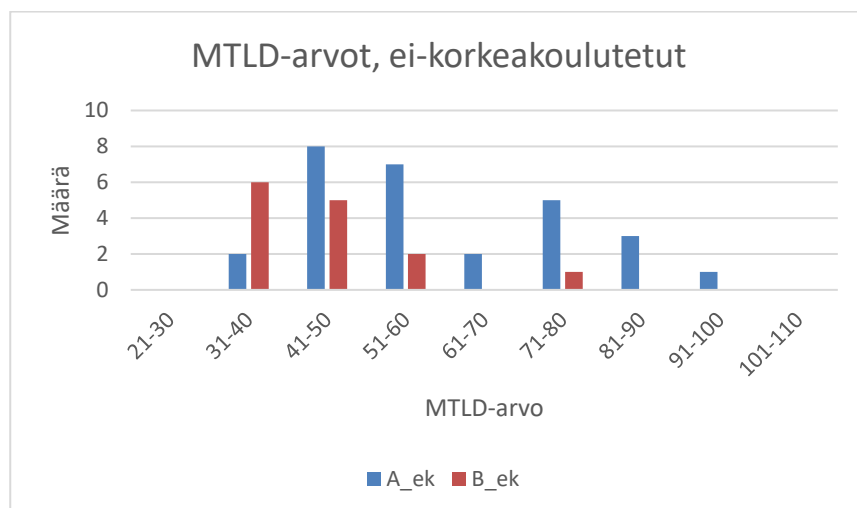
Kuvioissa 4 ja 5 olen jakanut ryhmät koulutuksen mukaan, jotta saisin tarkemman kuvan, miten korkeakoulutus vaikuttaa leksikaaliseen diversiteettiin. Kuvioista 4 näkyy, että eniten korkeita MTLD-arvoja on ryhmän A korkeakoulutetuilla ja yksi poikkeuksellisen suuri arvo ryhmässä B. Ryhmässä A seitsemällä informantilla MTLD-arvo on yli 70 ja ryhmässä B vain yhdellä.

Kuvio 4. A ja B ryhmien korkeakoulutettujen MTLD-arvojen jakauma



Samantapainen jaottelu huomattiin myös ei-korkeakoulutettujen kesken (ks. Kuvio 5), jossa ryhmän A yhdeksällä informantilla MTLD-arvo on yli 70 ja vain yhdellä ryhmässä B.

Kuvio 5. A ja B ryhmien ei-korkeakoulutettujen MTLD-arvojen jakauma



## MTLD ja ikä

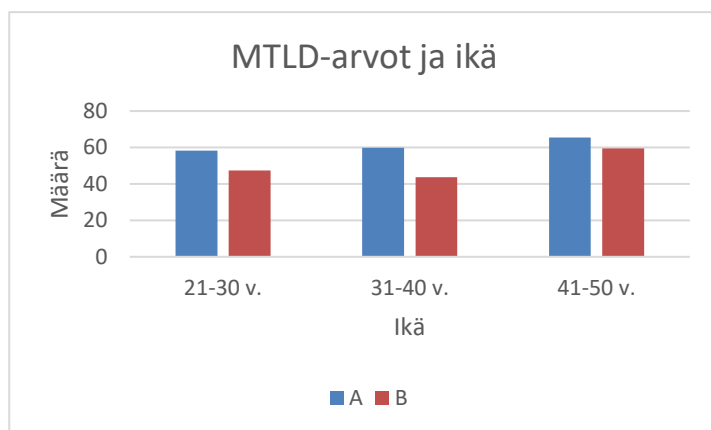
Ryhmän A informantit ovat jokaisessa ikäryhmässä saaneet korkeammat arvot kuin ryhmän B informantit.

Taulukko 7. A ja B ikäryhmien MTLD-keskiarvot ja mediaaniarvot

Ikä	Ryhmä A	Ryhmä B	Ryhmä A	Ryhmä B
	MTLD-keskiarvo		MTLD-mediaani	
21–30 v.	<b>58,35</b>	47,33	56,03	44,43
31–40 v.	<b>59,89</b>	43,78	55,27	42,96
41–50 v.	<b>65,49</b>	59,45	67,89	<b>47,42</b>

Taulukosta 7 näkyy, että töissä käyvien MTLD-arvot ovat väliltä 58–65, kun taas töitä vailla olevien väliltä 47–59. Vaikka ryhmän B 41–50-vuotiaiden ikäryhmäläisten MTLD-keskiarvo on huomattavasti korkeampi kuin muiden kahden ikäryhmän, se ei välttämättä anna oikeaa kuvaa ryhmän henkilöiden todellisista MTLD-arvoista. Ryhmän A ikäryhmässä on 10 informanttia, kun taas ryhmässä B on vain neljä informanttia. Yhden informantin MTLD-arvo näistä neljästä informantista on poikkeuksellisen korkea (101,87) ottaen huomioon sen, että B-ryhmäläisten mediaani on 47,42. Tämä yksi arvo nostaa koko ikäryhmän keskiarvoa. Se selittää, miksi ryhmän B 41–50-vuotiaiden ikäryhmän arvo on niin lähellä ryhmän A 41–50-vuotiaiden arvoa (ks. Kuvio 6). Epätasaisen informanttien määrän takia ikäryhmien vertailussa kannattaa tässä tutkimuksessa tarkastella myös mediaaniarvoja.

*Kuvio 6. A ja B ryhmien MTLD-arvojen jakautuminen iän mukaan*



## **MTLD ja tekstien pituus**

*Taulukko 8. Tekstipituus ja MTLD-keskiarvot*

Pituus	A (keskiarvo)	B (keskiarvo)
0...135,5	56,19	48,52
135,51...∞	64,19	46,47

Taulukossa 8 vertaan kahden ryhmän MTLD-keskiarvoja ja tekstien pituuksia, koska halusin tarkistaa, onko leksikaaliseen diversiteettiin ja sanojen määrän välillä yhteys ja jos on, niin millainen. Kriteeriksi valitsin molemmille ryhmille yhteisen tekstien pituuksien keskiarvon, joka on 135,5. Olen laskenut MTLD-keskiarvon niille informanteille, joiden tekstien pituudet ovat alle 135,5 ja niille, jotka ovat kirjoittaneet keskiarvoa pidemmät tekstit. Totesin, että ryhmässä A leksikaalinen diversiteetti kasvaa tekstin sanamäärän kanssa. Mielenkiintoista on, että ryhmästä B kävi ilmi, että kun teksti pitenee, MTLD-arvo laskee. Se tarkoittaa sitä, että erilaisten sanojen määrä tekstissä vähenee, eli samoja sanoja toistetaan useammin. Ryhmässä A tätä ei huomattu.

## Pienimmät ja suurimmat MTLT-arvot

Taulukko 9. Ryhmien minimi ja maksimi MTLT-arvot, -keskiarvot ja -mediaanit

Ryhmä	min. MTLT-arvo	maks. MTLT-arvo	MTLT-keskiarvo	MTLT-Mediaani
A_kk <sup>1</sup>	30,22	99,08	59,11	55,27
B_kk <sup>1</sup>	35,43	101,87	50,13	44,43
A_ek <sup>2</sup>	37,12	96,41	60,93	59,39
B_ek <sup>2</sup>	32,01	75,00	44,74	42,64

kk<sup>1</sup> – korkeakoulutetut, ek<sup>2</sup> – ei-kekeakoulutetut

Taulukossa 9 olen jakanut vertailtavat ryhmät minimi ja maksimi MTLT-arvojen mukaan. Ryhmän A kekeakoulutettujen mediaani on 55,27 ja ei-kekeakoulutettujen 59,39. Ryhmän B kekeakoulutettujen mediaani on 44,43 ja ei-kekeakoulutettujen 42,64. Numeraalisen datan perusteella voidaan sanoa, että ryhmien sisäiset erot ovat pieniä. Kuitenkin jos vertaamme ryhmän A ja B kekeakoulutettujen mediaaniarvoja, voimme huomata, että ryhmän A arvo (55,27) on suurempi kuin ryhmän B (44,43). Sama jaottelu näkyy myös ryhmän A ja B ei-kekeakoulutettujen välillä, joiden vastaavat arvot ovat 59,39 ja 42,64.

## Kirjoitustekstien esimerkit

Tässä esitän esimerkkitekstejä, jotka kuvaavat minimi ja maksimi MTLT-arvoja. Nämä ovat alkuperäisiä tekstejä, joita ei ole korjattu. Tekstit ovat anonyymejä ja tekstissä mainittu nimi *M. Solki* ei ole kenenkään informanteista oikea nimi, vaan se oli annettu tehtävänannossa ja se toistuu monissa muissa informanttien kirjoittamissa teksteissä. Sama koskee paikannimeä *Solkila*.

**A\_kk min. 30,22** tekstin pituus on 80

*Minä asun Solkilalla . Se ei ole iso kaupunki , mutta siin' on rakennettu kauppakeskuksia jo tarpeeksi! Meillä on kaksi kauppakeskuksia ; „ Hansa " ja „ Sokos ". Paitsi niitä on olemassa erilaisia pieniä kauppvoja. Se on riittä , ei tarvi rakentaa enempaa !  
Minä olen sitä mieltä , koska*

- 1. minä en kerki käydä kaikessa kaupoissa.*
- 2. On vaikea päättää mitä pitää ostaa .*
- 3. Ja rahaa ei riittää mihinkään ! Minun palkka on vain 1.000 euroa kuukaudessa!*

**A\_kk max. 99,08** tekstin pituus on 139

*Haluaisin kirjoittaa koulujen kesäloma ajalta. Olen ammattitani entinen ala-aste opettaja. Olen työskennellyt Venäjällä luokka opettajana monta vuotta ja olen tottunut siihen, että kesäloma alkaa kesäkuussa ja loppuu elokuussa. On riittävästi aikaa leppäämiseen ja matkustamiseen. Sekä lapsi, että vanhemmat saa levätä pitkästä opiskelujaksosta. Varsinkin ala-asteella vanhemmat joutuvat paljon auttamaan oma lastansa läksyjen tekemisessä. Saisimme rentoutua mekin (puhun*



vanhempien puolesta, koska olen itse kahden lapsen äiti). Sitten lisäksi Suomalainen kesä on niin lyhyt, että kannatta hyödyntää täysillä vapaa aikaa. Haluttaisi käydä ulkomaanreisussa, sukulaisten luona, mummolla, tapaamassa serkkuja ja paljon muuta kin kesäharrastuksija olisi kiva kokeilla. Mutta lapsi pitäisi laittaa kouluun jo 15 elokuuta. Kaikki asiat jäi kesken, ei keretty paljon mitään.

Olisin toivonut kesälomän aloittamaan 1 päivä kesäkuuta ja uudeilla voimilla aloittaa uutta kouluvuotta 1 syyskuuta.

Tämä vain minun oma mielipiteeni, voi olla monet vanhemmat ja lapset kaipaava samaa.

M. Solki.

**B\_kk min 35,43** tekstin pituus on 79

Haluaisin kirjoittaa vuokra-asuntosta ja omistusasuntosta. Pidän omistusasuntoa. Mutta usein se on kallis. Talon ostaminen on vaikea perheelle. Jos perheenjäsenillä on töitä, he voivat ottaa rahaa pankista. Se on parempi, kuin asunnon vuokra. Omistusasunnossa ihmiset voivat ostaa hyvää kalusteita, tehdä remontti ja tehdä puutarhan talon ympärille. Ihminen on talon omistaja. Vuokra-asunnossa ihminen ei teke mitään. Sinun täytyy maksaa rahaa ja et ole omistaja. Et voi tehdä sen, että haluat tehdä.

Minusta, omistusasunto on parempi, mutta kallimpi, Ihmisen täytyy valintaa itse.

**B\_kk max. 101,87** tekstin pituus on 42

Hei ! Haluaisin ilmoittaa oman mielipiteen luonnon ja ympäristön säilymisen. Luulen , että edes tavallisia ihmisiä voisivat vaikuttaa. Meidän aikana tämä asia on erittäin tärkeä. Esimerkiksi uudet energiasäästölamput sisältävät vaarallisia aineita. Niitä pitää laittaa oikealle paikalle , ei saa kadulle tai metsälle.

**A\_ek min. 37,12** tekstin pituus on 66

Paras hoitopaikka pienelle lapselle - oma koti.

Jos naine hoitaa lapsia kotona, tämä on rauhallinen paikka, lämmit ruoka. Naisille on aika leikia pihalla, jos sairas lapset antaa lääke. Pieni lapset kotona iloisempi, tervelisempi. Lapset oppi kotona paljon myyt asiat. Toinen kieli, maalaus, musikista. Lapset voi soittaa piano, kunnella radiosta lasten ohjelma. Hän voi katso televisiosta myös lasten ohjelmat. Iiksi minä ajatelen kotona on parempi, kuin päiväkodissä.

**A\_ek max. 96,41** tekstin pituus on 141

Monet meistä mietii tarkkään ennen ensiasunnon ostamista. Sehän on selvä: asuntolaina on todella iso ja maksuaika on pitkä. Rahaa matkustamiseen tai huiiin ei jää paljon mitään. Vuokralla asuminen olisi edullisempi. Mutta jos katsotaan pidemmän päälle, asiat ovat toisin. Vuokrankorotukset tulevat viime aikoina lähes joka vuosi, ja pian te maksatte saman verran kuukaudessa kuin omistusasunnon omistaja maksaa lyhennyksen ja koron pankille. Eli ero vain hoitovastikkeen maksun verran. Toinen omistusasunnon puolesta puhuva asia on meidän eläkevuodet. Eläke on n 40% palkka pienempi, ja vuokran maksamiseen siitä lahtee suurin osa rahasta. Omistusasunnon omistaja on jo maksanut asuntovelkensä pois, ja kuukausimaksuksi jää melko pieni hoitovastike. Ja kolmas, ja monista naisista paras argumentti, on se, että omistusasunto saa rakentaa ja sisustaa oman maun mukaan (jos tietenkin kukkaro sallii). Vuokra-asunnoissa on

*valmiiksi sisäänrakennettu keittio, yksinkertaisesti kaakeloitu kylpyhuone ja muovilattiat. Mitäs te nyt vastaatte kysymykseen vuokra-asunto vai omistusasunto?*

**B\_ek min. 32,01** tekstin pituus on 50

*Stop!?*

*Miksi kaikilla on niin kiire? Ajatelen, että kaikki halusivat ehtiä kaikki, mutta ei olle sitä mahdollisuutta. Sama kun et voi saada kaikkein rahaa elämäänsi.*

*Ihmisen elämä on liian lyhyt. Ei voi kuka pysähtyy aikaa, se menee aina yhä nopeampi ja nopeampi. Ihmiset ymmärsivät se ja heillä on aina kiire.*

**B\_ek max. 75,00** tekstin pituus on 55

*Minun mielestäni koulujen kesälomat ovat väärään aikaan.*

*Esim. Suomenkoulussa kesälomat kestää 2,5 kuukautta, mutta toisessa maassa kesäloma 3 kuukautta. Se on tosi mukavaa asiaa opiskelijoille. Kesälommalla täytyy uida, ottaa aurinkoa, Mutta suomenkoululaiset ei voi rentoutua normalisti, koska koululaiselle pakko mennä kouluun elokuun alussa. Heille täytyy istua koko päivä luokassa ja opiskella, mutta ulkona aurinko paistaa,*

Keskeinen ero ryhmien A ja B teksteissä on se, että ryhmän A informanttien tekstit ovat pidempiä kuin ryhmän B informanttien. Pidemmässä tekstissä voi helposti käydä niin, että kirjoittaja alkaa toistaa sanoja ja silloin MTLD-arvot alkavat laskea. Kuitenkin voimme nähdä tekstiesimerkistä A\_kk, jonka MTLD-arvo on 99,08, ja tekstiesimerkistä A\_ek (MTLD-arvo 96,41), että pitkästä tekstistä huolimatta MTLD-arvo voi pysyä korkeana. On mielenkiintoista, että ryhmässä B ei- korkeakoulutettujen minimi ja maksimi MTLD-arvojen välillä on suuri ero, mutta tekstien pituudet eivät juuri eroa. Poikkeuksellisesti lyhyessä B\_kk tekstissä MTLD-arvo on korkea.

#### 4.3 Lineaarinen regressioanalyysi

Käytän tässä tutkimuksessa tilastollisena menetelmänä monimuuttujaista lineaarista regressiota. Tämän menetelmän tehtävä on tarkastella, miten minun valitsemani muuttujat vaikuttavat toisiinsa, eli millainen lineaarinen yhteys niiden välillä on. Regressiomenetelmä on matemaattinen malli, joka kuvaa muuttujien välisiä yhteyksiä. Mallin rakentamiseen tarvitaan y-muuttuja ja x-muuttuja, jotta saamme kuvan näiden muuttujien välisistä suhteista. Tämän menetelmän avulla y-muuttujan arvojen vaihtelu pystytään selittämään x-muuttujalla. Kun tiedetään x-pistemäärä, voimme ennustaa y-pistemäärän ja päinvastoin. Tässä x-muuttuja on selittäjä ja y-muuttuja on selitettävä. Malliin voi lisätä tai poistaa muuttujia kerrallaan. Monimuuttujainen regressiomalli toimii samalla tavalla kuin

yhden selittäjän regressiomallit, ja sen regressiosuoran yhtälö on vastaavalainen kuin yhden selittäjän regressiomalli. Usean muuttujan regressiomallia pidetään luotettavimpana, koska sen tulos on tarkempi. Monimuuttujaiseen regressiomalliin otetaan useita x-muuttujia ja regressiosuoran yhtälö muodostetaan alla mainitulla tavalla:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k,$$

jossa  $\hat{y}$  = y:n ennustettu arvo,  $\beta_0$  = vakiotermi,  $\beta_1, \dots, \beta_k$  = regressiokertoimet,  $x_1, \dots, x_k$  = x-muuttujien arvot. (Nummenmaa 2009.)

Käyttämäni aineiston ja alaluvussa 2.2 esittelemieni sanaston oppimiseen vaikuttavien tekijöiden perusteella päätin valita regressioanalyysiin neljä muuttujaa: työllistyminen, koulutus, ikä ja tekstin pituus. Ensiksi rakensin matemaattisen mallin, jossa käytän kaikkia näitä neljää muuttujaa, ja sen jälkeen jokaisessa erillisessä mallissa poistan yhden muuttujan kerrallaan. Tarkoitus on löytää paras malli, joka sopisi parhaiten tähän aineistoon. Seuraavassa esitän kaikki nämä tehdyt mallit. Malleissa y on MTLN-arvo ja x-muuttujat ovat työssä käyminen, koulutus, ikä ja tekstin pituus. Tein taulukon jokaisesta mallista, johon merkitsin muuttujien yksittäiset p-arvot. Taulukon jälkeen merkitsin  $R^2$ -arvon (mallin selitysaste), korjatun  $R^2$ -arvon (korjattu selitysaste) ja mallin p-arvon. Mallin selitysaste ( $R^2$ ) osoittaa, kuinka paljon y-muuttujan (MTLN-arvo) vaihtelusta voidaan selittää mallin avulla. Kun tehdään useita malleja ”korjattu selitysaste” on luotettavampi, koska se ottaa huomioon muuttujien lukumäärän ja kuvaa mallin sopivuutta siihen aineistoon, josta otos on otettu.

### **Malli nro 1: Työllistyminen + koulutus+ ikä+ pituus**

Ensimmäinen malli koostuu kaikista mahdollisista muuttujista: työllistyminen, koulutus, ikä ja tekstin pituus. (F (4,75) = 2.704, p<0.05)

<b>Muuttuja</b>	<b>yksittäiset p-arvot</b>
Työllistyminen	0.00322
Koulutus	0.82077
Ikä	0.52850
Tekstin pituus	0.92934

$R^2 = 0.126$ , korjattu  $R^2 = 0.07942$ , p = 0.03664

### **Malli nro 2: Työllistyminen + koulutus + ikä + ~~pituus~~**

Toisessa mallissa poistin tekstin pituuden muuttujista. ( $F(3,76) = 3.65$ ,  $p < 0,05$ )

Muuttuja	yksittäiset p-arvot
Työllistyminen	0.00303
Koulutus	0.82526
Ikä	0.52985

$R^2 = 0.1259$ , korjattu  $R^2 = 0.09144$ ,  $p = 0.01621$

### **Malli nro 3: Työllistyminen + koulutus + ~~ikä~~ + ~~pituus~~**

Kolmannessa mallissa poistin iän. ( $F(3,76) = 3.499$ ,  $p < 0,05$ )

Muuttuja	yksittäiset p-arvot
Työllistyminen	0.00182
Koulutus	0.83022
Pituus	0.99467

$R^2 = 0.1214$ , korjattu  $R^2 = 0.08668$ ,  $p = 0.01946$

### **Malli nro 4: Työllistyminen + ~~koulutus~~ + ikä + ~~pituus~~**

Neljännessä mallissa otin pois koulutuksen. ( $F(3,76) = 3.633$ ,  $p < 0,05$ )

Muuttuja	yksittäiset p-arvot
Työllistyminen	0.0031
Ikä	0.5289
Pituus	0.9453

$R^2 = 0.1254$ , korjattu  $R^2 = 0.09091$ ,  $p = 0.01654$

### **Malli nro 5: ~~Työllistyminen~~ + koulutus + ikä + ~~pituus~~**

Viidennessä mallissa tarkastelen mallia ilman työllistymistä muuttujana. ( $F(3,76) = 0.4656$ ,  $p > 0,05$ )

Muuttuja	yksittäiset p-arvot
Koulutus	0.977452
Ikä	0.249201
Pituus	0.957729

$R^2 = 0.01805$ , korjattu  $R^2 = -0.02071$ ,  $p = 0.7072$

### **Malli nro 6: Työllistyminen + Koulutus + ikä + pituus**

Kuudennesta mallista jätin pois sekä tekstin pituuden että iän. ( $F(2,77) = 5,318$ ,  $p < 0,05$ )

<b>Muuttuja</b>	<b>yksittäiset p-arvot</b>
Työllistyminen	0.00165
Koulutus	0.82891

$R^2 = 0.1214$ , korjattu  $R^2 = 0.09854$ , p-arvo = 0.006866

### **Malli nro 7: Työllistyminen**

Seitsemännessä mallissa tarkastelen vain yhtä muuttujaa: työllistymistä. ( $F(1,78) = 10.72$ ,  $p < 0,05$ )

<b>Muuttuja</b>	<b>p-arvo</b>
Työllistyminen	0.00158

$R^2 = 0.1208$ , korjattu  $R^2 = 0.1096$ ,  $p = 0.001582$

Mallin tulokset osoittavat, että valitsemallani muuttujalla voidaan selittää vain 11 % MTLA-arvojen vaihtelusta, eli mallin selitysaste ei ole kovin korkea ( $R^2 = 0.1096$ ). Malli osoittaa, että on myös muita tekijöitä, jotka vaikuttavat MTLA-arvoon.

Tämän tutkimuksen kannalta paras ja tärkein malli on yhden muuttujan malli, koska p-arvo (0.001582) on pienempi kuin muiden mallien p-arvo ja myös pienempi kuin 0,05, joka on tilastollisen merkittävyyden raja (ks. Taulukko 10).

*Taulukko 10. Monimuuttujainen lineaaristen regressiomallien vertailu*

	<b><math>R^2</math></b>	<b>Korjattu <math>R^2</math></b>	<b>p-arvo</b>
<b>Malli 1</b>	0.126	0.07942	0.03664
<b>Malli 2</b>	0.1259	0.09144	0.01621
<b>Malli 3</b>	0.1214	0.08668	0.01946
<b>Malli 4</b>	0.1254	0.09091	0.01654
<b>Malli 5</b>	0.01805	-0.02071	0.7072
<b>Malli 6</b>	0.1214	0.09854	0.006866
<b>Malli 7</b>	0.1208	0.1096	<b>0.001582</b>

## 5 Yhteenveto

### 5.1 Johtopäätökset

Kasvava maahanmuutto työn perässä Suomeen lisää tarvetta tutkia S2-oppijoiden kieltä ja sen kehittymisen mahdollisuuksia. Oppijat ovat erilaisia luonteeltaan, oppimistottumuksiltaan, motivaatioltaan ja asenteeltaan. Kaikki nämä asiat vaikuttavat tavalla tai toisella oppimiseen. Tässä tutkimuksessa tarkastelin, miten tekijät, kuten työllistyminen, koulutus ja ikä mahdollisesti voivat vaikuttaa sanaston osaamiseen. Työympäristö voi tarjota affordansseja oppijalle, mutta oppijan pitäisi olla aktiivinen, jotta hän pystyisi hyödyntämään näitä tarjoumia. Pedagogisesta näkökulmasta katsoen opettajan kannattaa ohjata oppilaita etsimään affordansseja luokan ulkopuolelta. Oppiminen voi tapahtua ihan missä vaan, eikä se ole kiinni paikasta. Uutta voi oppia työpaikalla tai puhumalla ystävän kanssa. Formaali oppiminen ei voi tarjota oppijalle kaikkea ja taata riittävää harjoitusta. Tällöin informaali oppiminen on yksi vaihtoehto, miten oppija voi lisätä omalle oppimiselleen vuorovaikutusta ja harjoituksia. Oppija usein kuuluu enemmän kuin yhteen oppimiskontekstiin ja kokee muutamaa eri oppimismuotoa.

Tutkimukseni tavoitteena oli selvittää, miten kohdemaan ympäristö vaikuttaa oppijan sanavaraston laajuuteen. Tämän tutkimuksen kohderyhmäksi valitsin venäjänkieliset, jotka ovat suorittaneet suomen kielen keskitason YKI-tutkinnon. Informantit jaoin kahteen ryhmään kielenkäyttötilanteen mukaan. Ensimmäisen ryhmän (A) informantit ovat täysin altistuneita suomen kielelle, koska he käyttävät suomen kieltä lähes päivittäin kotonaan, töissä, vapaa-aikanaan ja opiskeluissaan. Toisen ryhmän informantit käyttävät suomen kieltä ainoastaan vapaa-aikana tai opiskelussa. Suurin ero ryhmien välillä on se, että ryhmän B informantit eivät käytä kohdekieltä töissä. Se vähentää altistumismahdollisuuksien määrää suomen kielelle. Tutkittavaksi valitsin YKI-tutkinnon kirjoittamisen suorituksia, joiden perusteella tutkin leksikaalista diversiteettiä. Sen mittarina käytin MTLN-arvojen laskemista, jossa suurempi MTLN-arvo osoittaa, että kielenkäyttäjän sanasto on laajempi.

Tässä pro gradu -tutkimuksessani pyrin vastamaan siihen, miten työssä käyminen ja korkeakoulutus vaikuttavat sanaston laajuuteen. Hypoteesini oli se, että henkilöiden, jotka käyvät töissä, sanavarasto on laajempi kuin niiden, jotka ovat työttömiä. Oletukseni oli se, että ympäristön tarjoamat oppimis- ja käyttömahdollisuudet, kuten radion kuuntelu, televisio-ohjelmien seuraaminen ja sosiaaliset kontaktit suomalaisten kanssa vaikuttavat positiivisesti sanaston osaamiseen. Analyysiosassa tarkastelin aineiston MTLN-arvoja työllisyys tilanteen, koulutuksen ja iän perusteella. Totesin, että

eniten leksikaaliseen diversiteettiin vaikuttaa työssä käyminen, jossa MTLD-keskiarvot ryhmän A ja B kesken olivat 60,11 ja 47,53. Tämä vahvistaa hypoteesini siitä, että työympäristö auttaa laajentamaan henkilön sanavarastoa. Kuten Sunin tutkimus (2010) osoitti, sosiaalinen vuorovaikutus ihmisten kanssa kohdekielellä on merkittävä tekijä kielen oppimisessa, ja uusia sanoja voi oppia työpaikalla ja asiakkailta. Työllisyystilanteen vaikutus leksikaaliseen diversiteettiin todettiin olevan tilastollisesti merkitsevä. Tämän tutkimuksen tuloksien perusteella voidaan todeta, että töissä käyminen vaikuttaa leksikaaliseen diversiteettiin positiivisesti, eli leksikaalisen diversiteetin arvo kasvaa.

Toinen hypoteesini oli se, että korkeakoulutettujen sanasto on laajempi. Koulutuksella voi olla merkitystä siinä mielessä, onko henkilöllä korkeakoulutus vai alemman tason koulutus. Ryhmän A, eli työssä käyvien, informanttien osalta ei koulutus näytä vaikuttavan leksikaaliseen diversiteettiin. Kun kielenoppija taas ei käy töissä, näyttää koulutustaustalla olevan vaikutus MTLD-arvoon. Saatujen MTLD-arvojen perusteella kävi ilmi, että ryhmän B korkeakoulutettujen informanttien MTLD-keskiarvo on 50,13 ja ei-ke korkeakoulutettujen 44,74. Nämä tulokset osoittavat, että koulutuksella ja työllistymistilanteella on yhteys sanaston laajuuteen. Korkeakoulutetuilla, jotka tutkinnon suorituspäivänä ovat olleet työttömiä, on laajempi sanasto kuin ei-ke korkeakoulutetuilla työtä vailla olevilla.

Tutkimuksen tilastollisena menetelmänä käytin monimuuttujaista lineaarista regressiota. Rakensin matemaattisen mallin, jossa muuttujina käytin työllistymistä, koulutusta, ikää ja tekstin pituutta. Sen jälkeen rakensin muut mallit siten, että poistin joka kerta yhden muuttujan alkuperäisestä mallista. Tällä menetelmällä rakensin yhteensä seitsemän mallia. Parhaiten tähän tutkimukseen sopiva malli on yhden muuttujan malli. Yhden muuttujan (työssä käyminen) malli osoittaa, että on 99,8 %:n todennäköisyys, että työllistyminen vaikuttaa MTLD-arvoon ( $p=0.00158$ ). Kuitenkin vain 11 % ( $R^2=0.1096$ ) sanaston monimuotoisuuden vaihtelusta voidaan selittää rakennetun mallin avulla ja osa aineistossa havaitusta vaihtelusta jää selittämättä. Tämä tarkoittaa sitä, että on olemassa myös muita tekijöitä, jotka vaikuttavat sanaston kasvuun, mutta joita emme vielä tunne. Tämän tutkimuksen osalta voidaan sanoa, että työssä käymisellä on yhteys laajempaan sanastoon. Kuitenkaan tämän tutkimuksen perusteella ei voida sanoa, mitkä kaikki muut tekijät vaikuttavat sanaston laajuuteen. Tämän vuoksi tarvitaan lisätutkimusta, jotta voimme saada tietoa tekijöistä, jotka jäävät tämän mallin perusteella selittämättä.

## 5.2 Pohdinta

Tutkimuksen alkuperäinen ajatus oli vertailla keskenään niitä osallistujia, jotka ovat opiskelleet kieltä ulkomailla ja niitä, jotka opiskelevat suomea toisena kielenä Suomessa. Taustatietolomakkeen mukaan YKI-testin suorittaneet henkilöt ovat opiskelleet suomea monissa eri konteksteissa ja heidän opiskelujensa kestot ovat vaihdelleet. Informanttien taustaa ja kielenkäytön mahdollisuuksia olen arvioinut taustatietolomakkeen perusteella. Kuitenkaan emme voi tietää, miten huolellisesti informantit ovat täyttäneet taustatietolomakkeet ja pitävätkö annetut tiedot paikkansa. Tämän tutkimuksen kannalta taustatietolomake on hyvin suppea, koska sen perusteella ei saa tarkkoja tietoja siitä, millaisten aktiviteettien parissa informantit ovat olleet tekemisissä suomen kielen kanssa. Tätä tutkimusta tehdessäni kaipasin tietoa esimerkiksi siitä, millainen sosiaalinen vuorovaikutus informanteilla on, katsovatko he suomalaista televisiota, seuraavatko he sosiaalista mediaa ja missä määrin. Uskon, että samalaisia havaintoja ovat tehneet myös YKI-korpuksen laatijat, koska aineiston, joka on kerätty vuoden 2011 jälkeen, taustatietolomake on uusittu ja tutkintokielen käyttämisen kohtaa on laajennettu. Tämä tutkimusaineisto olisi ollut vertailukelpoisempi, jos siinä olisi ollut tarpeeksi paljon kirjoittamisen suorituksia samasta aiheesta. Silloin sanaston osaamiseen liittyvät tulokset heijastaisivat paremmin informanttien leksikaalista osaamista. Tässä tutkimuksessa tutkittiin vain yhtä jokaisen informantin kirjoittamaa tekstiä työn rajatun laajuuden takia. Voi olla, että lisätekstien käyttö olisi antanut varmemman kuvan informanttien sanaston osaamisesta.

Tulisi muistaa, että YKI-tutkintotilanne ei ole luonnollinen, koska se tehdään koetilanteessa. Tutkinnon suorittaja ei testin aikana välttämättä pysty näyttämään parasta osaamistaan esimerkiksi psykologisista tai fyysisistä syistä tai testiin määritettyjen teemojen takia. Voi hyvin olla, että suorittajaa jännitti tai hän oli nukkunut huonosti ennen suorituspäivää tai hänellä muuten oli huono päivä, ja kaikki nämä asiat voivat johtaa huonompaan suoritukseen. Suoritus voi myös riippua tekstin aiheesta tai ajan paineesta (testissä on rajoitettu aika suorittaa tehtävät). Yksi aihe voi olla tutumpi kuin toinen riippuen ihmisen kokemuksista, iästä ja taustasta. Tässä tutkimuksessa tutkittiin produktiivisen sanaston osaamista, mutta pitäisi ottaa huomioon, että reseptiivinen sanasto on usein suurempi. Tämä tarkoittaa sitä, että henkilö osaa enemmän sanoja kuin hän on kykenevä tuottamaan tiettyssä tilanteessa esimerkiksi testissä.

On myös muita tekijöitä, jotka mahdollisesti vaikuttavat informanttien sanaston osaamiseen. Tulevaisuuden tutkimuksissa olisi mielenkiintoista selvittää, mitä nämä tekijät ovat. YKI-korpus on jatkuvasti kasvava korpus, koska aineistoa lisätään sitä mukaan, kun ihmiset suorittavat testejä. Tämä mahdollistaisi kattavampia lisätutkimuksia aiheesta tulevaisuudessa.



## Lähteet

### Aineslähteet

YKI-korpus. Yleisten kielitutkintojen testiaineistosta koostettu korpus. Jyväskylän yliopiston Soveltavan kielentutkimuksen keskus. Saatavissa: <http://yki-korpus.jyu.fi/>

### Muut lähteet

Ahola, S. – Lammervo, T. 2013. *Venäjää äidinkielenään puhuvien maahanmuuttajien suomen kielen taito yleisten kielitutkintojen valossa*. Siirtolaisuus Migration, 40:1. [Viitattu 21.1.2019.] Saatavissa: <https://siirtolaisuus-migration.journal.fi/issue/view/6161/884>

Aikuisten maahanmuuttajien kotoutumiskoulutuksen opetussuunnitelman perusteet 2012. [Viitattu 20.4.2020.] Saatavissa: [https://www.oph.fi/sites/default/files/documents/139342\\_aikuisten\\_maahanmuuttajien\\_kotoutumiskoulutuksen\\_opetussuunnitelman\\_perusteet\\_2012.pdf](https://www.oph.fi/sites/default/files/documents/139342_aikuisten_maahanmuuttajien_kotoutumiskoulutuksen_opetussuunnitelman_perusteet_2012.pdf)

Bulté, Bram – Housen, Alex – Pierrard, Michel – Van Daele, Siska 2008. *Investigating Lexical Proficiency Development over Time – the Case of Dutch-Speaking Learners of French in Brussels*. Journal of French Language Studies 18:3, s. 277–298. [Viitattu 27.5.2020.] Saatavissa: <https://doi.org/10.1017/S0959269508003451>

COE = Council of Europe. [Viitattu 5.2.2020.] Saatavissa: <https://www.coe.int/fi/web/lang-migrants/cefr-and-profiles>

Collentine, Joseph – Freed, Barbara F. 2004. *Learning context and its effects on second language acquisition: Introduction*. Studies in Second Language Acquisition 26:2, s. 153–171. [Viitattu 5.2.2020.] Saatavissa: <https://doi.org/10.1017/S0272263104262015>

Collentine, Joseph 2004. *The effects of learning contexts on morphosyntactic and lexical development*. Studies in Second Language Acquisition 26:2, s. 227–248. [Viitattu 27.5.2020.] Saatavissa: <https://doi.org/10.1017/S0272263104262040>

Ellis, R. 2008. *The study of second language acquisition*. 2nd edition. Oxford

EVK = Eurooppalainen viitekehys. [Viitattu 21.1.2019.] Saatavissa: <http://www02.oph.fi/ops/taitotasoasteikko.pdf>

Foster, P. 2009. *Lexical diversity and native-like selection: The bonus of studying abroad*. Teoksessa B. Richards et al. (Eds.), *Vocabulary studies in first and second language acquisition*, s. 91–106. Hampshire, UK: Palgrave Macmillan.

Freed, B. – Segalowitz, N. – Dewey, D. P. 2004. *Context of learning and second language fluency in French: Comparing Regular Classroom, Study Abroad, and Intensive Domestic Immersion Programs*. Studies in Second Language Acquisition, 26:2, s. 275–301. [Viitattu 5.2.2020.] Saatavissa: <https://doi.org/10.1017/S0272263104262064>

Gibson, James J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin,

González-Fernández Beatriz – Schmitt Norbert 2017. *Vocabulary Acquisition*. Teoksessa Loewen, Shawn – Masatoshi, Sato. *The Routledge Handbook of Instructed Second Language Acquisition*. Milton Park, Abingdon, Oxon: Routledge.

Hakala, Jarkko 2018. *Ulkomailta tulleiden lääkäreiden sanasto laillistamiskuulustelussa*. Pro gradu -tutkielma. Tampereen yliopisto. [Viitattu 5.5.2020.] Saatavissa: <http://urn.fi/URN:NBN:fi:uta-201810122685>

- Honko, M.– Jarvis, S. – Vainio, S. 2019. *Lukijat sanaston monimuotoisuutta määrittämässä: Leksikaalisen diversiteetin tarkastelua määrällisen ja laadullisen tutkimuksen rajapinnassa*. Virittäjä, 123:1. [Verkkolehti.] Kotikielen Seura, Helsinki. [Viitattu 18.5.2020.] Saatavissa: <https://doi.org/10.23982/vir.59025>
- Honko, Mari 2013. *Alakouluikäisten leksikaalinen tieto ja taito: toisen sukupolven suomi ja S1-verrokki*. Väitöskirja. Tampereen yliopisto. [Viitattu 5.5.2020.] Saatavissa: <http://urn.fi/URN:ISBN:978-951-44-9251-8>
- Honko, Mari 2015. *Kohti vahvempaa sanastohallintaa – Ajatuksia sanastohallinnasta ja sanaston omaksumisen tukemisesta*. [Viitattu 1.6.2020.] Saatavissa: <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-huhtikuu-2015/kohti-vahvempaa-sanastohallintaa-ajatuksia-sanastohallinnasta-ja-sanaston-omaksumisen-tukemisesta>
- Honko, Mari 2017. *Sadutettu sanasto: puhutun kielen leksikaalinen diversiteetti arviointikohteena*. Soveltavan kielitieteen tutkimuksia AFinLA-e, 10. [Verkkolehti.] Suomen soveltavan kielitieteen yhdistys AFinLA ry, Jyväskylä. [Viitattu 27.1.2019.] Saatavissa: <https://journal.fi/afinla/article/view/73136>
- Huhta, A. – R. Hildén (toim.) 2016. *Kielitaidon arviointitutkimus 2000-luvun Suomessa*. Soveltavan kielitieteen tutkimuksia AFinLA-e 9, s. 3–26. [Verkkolehti.] Suomen soveltavan kielitieteen yhdistys AFinLA ry, Jyväskylä. [Viitattu 1.6.2020.] Saatavissa: <https://journal.fi/afinla/issue/view/4267>
- Integroituminen ammattiyhteisöihin Suomessa –Transforming Professional Integration 2011–2014. [Viitattu 27.1.2019.] Saatavissa: <https://www.jyu.fi/hytk/fi/laitokset/solki/tutkimus/hankkeet/isis>
- Itsearviointilokerikko [Viitattu 1.6.2020.] Saatavissa: <https://europa.eu/europass/fi/common-european-framework-reference>
- Jantunen, Jarmo 2015. *Oppimiskontekstin vaikutus oppijanpragmatiikkaan: astemääritteet leksikaalisina nallekarhuina*. Lähivertailuja 25 [Verkkolehti.] Eesti Rakenduslingvistika Ühing. [Viitattu 16.02.2020.] Saatavissa: <http://arhiiv.rakenduslingvistika.ee/ajakirjad/index.php/lahivordlusi/article/view/LV25.05>
- Jarvis, Scott 2013a. *Capturing the Diversity in Lexical Diversity*. Language Learning: A Journal of Research in Language Studies 63:1, s. 87–106. [Viitattu 16.02.2020.] Saatavissa: <https://doi.org/10.1111/j.1467-9922.2012.00739.x>
- Jarvis, Scott 2013b. *Chapter 1. Defining and measuring lexical diversity*. Teoksessa Scott Jarvis – Michael Daller. Vocabulary Knowledge: Human Ratings and Automated Measures [Viitattu 27.1.2019.] Saatavissa: <https://doi.org/10.1075/sibil.47.03ch1>
- Jarvis, Scott 2017. *Grounding lexical diversity in human judgments*. Language testing 34:4, s. 537–553. <https://doi.org/10.1177/0265532217710632>
- Jyväskylän koulutuskuntayhtymä [Viitattu 20.4.2020] Saatavissa: <https://peda.net/jao-ammattillinen/jamsa/yhteiset-opinnot/ovl21/p62/ajpl/mielipiteen-ilmaisu>
- Kansalaisuuslaki 359/2003 [Viitattu 27.1.2019.] Saatavissa: <http://www.finlex.fi/fi/laki/alkup/2003/20030359#Pidp446341840>
- Kauppalehti.fi *Maahanmuuttajanaisten kotoutus ontuu Suomessa - OECD antoi kolme suositusta* [Viitattu 27.1.2019.] Saatavissa: <https://www.kauppalehti.fi/uutiset/maahanmuuttajanaisten-kotoutus-ontuu-suomessa-oecd-antoi-kolme-suositusta/6554414b-c21f-3e1c-b209-721b08833abe>
- Kela, M. – Komppa, J. 2011. *Sairaanhoitajan työkieli – yleiskieltä vai ammattikieltä? Funktionaalinen näkökulma ammattikielen oppimiseen toisella kielellä*. Puhe ja Kieli, 31:4, s. 173–192 [Verkkolehti.] Helsingin yliopisto. [Viitattu 27.1.2019.] Saatavissa: <https://journal.fi/pk/article/view/4752>
- Keuleers, Emmanuel – Stevens, Michaël – Mandera, Paweł – Brysbaert, Marc 2015. *Word Knowledge in the Crowd: Measuring Vocabulary Size and Word Prevalence in a Massive Online Experiment*. The Quarterly Journal of Experimental Psychology: Megastudies, Crowdsourcing, and Large Datasets in Psycholinguistics 68:8, s. 1665–1692. [Viitattu 27.1.2019.] Saatavissa: <https://journals.sagepub.com/doi/10.1080/17470218.2015.1022560>

- Kielitaidon tasojen kuvausasteikko. [Viitattu 27.1.2019.] Saatavissa: [https://www.oph.fi/sites/default/files/documents/kielitaidon\\_tasojen\\_kuvausasteikko.pdf](https://www.oph.fi/sites/default/files/documents/kielitaidon_tasojen_kuvausasteikko.pdf)
- Kirjoittamisen arviointikriteerit. [Viitattu 28.1.2019.] Saatavissa: [https://www.jyu.fi/hytk/fi/laitokset/solki/yki/yleista/tietoakielitutkinnoista/Kirjoittamisen\\_arviointikriteerit.pdf](https://www.jyu.fi/hytk/fi/laitokset/solki/yki/yleista/tietoakielitutkinnoista/Kirjoittamisen_arviointikriteerit.pdf)
- Komppa J. – Jäppinen T. – Herva M. – Hämäläinen T. 2014. *Korkeakoulutuksen ammatilliset suomi toisena kielenä -viitekehikset*. AATOS-artikkelit 16. Metropolia Ammattikorkeakoulu. [Viitattu 8.6.2020] Saatavissa: <http://urn.fi/URN:ISBN:978-952-6690-12-4>
- Komppa, Johanna 2015. *Työnantajan odotukset, työntekijän vastuu ja työyhteisön tuki. Näkökulmia korkeakoulutettujen maahanmuuttajien ammatillisen suomen oppimiseen*. Soveltavan kielitieteen tutkimuksia AFinLA-e, 8. [Verkkolehti.] Suomen soveltavan kielitieteen yhdistys AFinLA ry, Jyväskylä. [Viitattu 27.1.2019.] Saatavissa: <https://journal.fi/afinla/article/view/53778>
- Laine-Leinonen Jaana 2013. *Koulukorpuksen leksikko: 1.-6.-luokkalaisten aktiivinen sanavarasto*. Pro gradu -tutkielma. Tampereen yliopisto. Saatavissa: <https://trepo.tuni.fi/handle/10024/84231>
- Laki kansalaisuuslain muuttamisesta 579/2011 [Viitattu 27.1.2019.] Saatavissa: <http://www.finlex.fi/fi/laki/alkup/2011/20110579>
- Lampinen, J. – M. Juntunen 2013. Oppijat – hankkeen tuotos. Informaali oppiminen. [Viitattu 28.3.2020] Saatavissa: <https://peda.net/hankkeet/oppijat/ovo/lahtokohdat/informaali>
- Lintunen, Pekka – Maarit Mutta – Sanna Pelttari 2017. *Profiling Language Learners in Hybrid Learning Contexts: Learners' Perceptions*. The EuroCALL Review 25:1, s. 61–75. [Viitattu 28.3.2020] Saatavissa: <https://doi.org/10.4995/eurocall.2017.7145>
- Llanes, Àngels – Carmen Muñoz 2013. *Age Effects in a Study Abroad Context: Children and Adults Studying Abroad and at Home*. Language Learning 63:1, s. 63–90. [Viitattu 28.3.2020] Saatavissa: <https://doi.org/10.1111/j.1467-9922.2012.00731.x>
- Magno, Carlo et al., 2009. *Assessing the Level of English Language Exposure of Taiwanese College Students in Taiwan and the Philippines*. Asian EFL Journal 11:1, s. 62–73. [Viitattu 28.3.2020] Saatavissa: [https://www.researchgate.net/publication/228225200\\_Assessing\\_the\\_Level\\_of\\_English\\_Language\\_Exposure\\_of\\_Taiwanese\\_College\\_Students\\_in\\_Taiwan\\_and\\_the\\_Philippines](https://www.researchgate.net/publication/228225200_Assessing_the_Level_of_English_Language_Exposure_of_Taiwanese_College_Students_in_Taiwan_and_the_Philippines)
- Malin, Essi 2012. *Suomi toisena kielenä -oppijoiden sanaston kehittyminen taitotasolta toiselle siirryttäessä*. Pro gradu -tutkielma. Jyväskylän yliopisto. Saatavissa: <https://jyx.jyu.fi/handle/123456789/40415>
- Masonen, Virpi 2013. *Leksikaalinen kompetenssi osana kielitaitoa*. Lähivertailuja 23. [Verkkolehti.] Eesti Rakenduslingvistika Ühing. [Viitattu 27.1.2019.] Saatavissa: <http://dx.doi.org/10.5128/LV23.08>
- McCarthy, Philip M – Jarvis, Scott 2010. *MTLD, Vocd-D, and HD-D: A Validation Study of Sophisticated Approaches to Lexical Diversity Assessment*. Behavior research methods 42:2, s. 381–392. Saatavissa: <https://doi.org/10.3758/BRM.42.2.381>
- Mutta, M. – Lintunen, P. – Pelttari, S. 2017. *Kielitaito ja informaalin oppimisen kontekstit: kielten opiskelijoiden ja yhteiskuntatieteilijöiden käsityksiä kielitaidosta ja teknologioiden hyödyntämisestä oppimisen tukena*. AFinLAn Vuosikirja, 75, 181–200. [Viitattu 11.8.2020.] Saatavissa: <https://journal.fi/afinlavk/article/view/60777>
- Nummenmaa, Lauri 2009. *Käyttätymistieteiden tilastolliset menetelmät*. 1. p., uud. laitos. Helsinki: Tammi.
- OPH = Opetushallituksen sivu. [Viitattu 21.1.2019.] Saatavissa: <https://www.oph.fi/fi/koulutus-ja-tutkinnot/yleinen-kielitutkinto-usein-kysytyt-kysymykset>

- Partanen, Maiju 2013. *Suomen kielen oppimisen mahdollisuudet ja työyhteisön tuki puhdistuspalvelualalla: afrikkalaisten maahanmuuttajien käsityksiä ja kokemuksia*. Teoksessa T. Keisanen, E. Kärkkäinen, M. Rauniomaa, P. Siitonen & M. Siromaa (toim.). Osallistumisen multimodaaliset diskurssit. AFinLA:n vuosikirja 2013. Suomen soveltavan kielitieteen yhdistyksen julkaisuja n:o 71. Jyväskylä: AFinLA, s. 55–76. [Viitattu 11.8.2020.] Saatavissa: <https://journal.fi/afinlavk/article/view/60043>
- Pietilä, Päivi – Lintunen, Pekka 2014. *Kuinka kieltä opitaan: opas vieraan kielen opettajalle ja opiskelijalle*. Helsinki: Gaudeamus.
- Pöyhönen, Sari; – Tarnanen, Mirja 2011. *Maahanmuuttajien suomen kielentaidon riittävyys ja työllistymisen mahdollisuudet*. Puhe ja Kieli, 4. [Verkkolehti.] Helsingin yliopisto. [Viitattu 28.1.2019.] Saatavissa: <https://journal.fi/pk/article/view/4750>
- Rynkänen, Tatjana 2015. *Venäjänkieliset maahanmuuttajat suomalaisissa työyhteisöissä*. Idäntutkimus 22:2, s. 85–86. [Verkkolehti.] Venäjän ja Itä-Euroopan tutkimuksen seura ry. [Viitattu 12.11.2020.] Saatavissa: <https://journal.fi/idantutkimus/article/view/78025>
- Sahradyan, Sonya 2015. *Miten maahanmuuttajat oppivat suomea formaalin, informaalin ja non-formaalin oppimisen kautta?* [Viitattu 28.2.2020] Saatavissa: <https://jyx.jyu.fi/handle/123456789/46816>
- Seilonen, Marja, – Suni, Minna 2016. *Ohjeita, Tietoa Ja Turvaa Kielen Keinoin. Ulkomailta Palkatut Sairaanhoidajat Ammatillista Suomen Kielen Taitoaan Osoittamassa*. Lähivertailuja 26, s. 450–480. [Verkkolehti.] Eesti Rakenduslingvistika Ühing. [Viitattu 16.7.2020.] Saatavissa: <http://dx.doi.org/10.5128/LV26.15>
- Solki = Soveltavan kielentutkimuksen keskus [Viitattu 27.1.2019.] Saatavissa: <https://www.jyu.fi/hytk/fi/laitokset/solki/yki/yleista/arviointi-ja-todistus>
- Straka, Milan – Straková, Jana 2017. *Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe*. In Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, 88–99. Vancouver, Canada: Association for Computational Linguistics. Saatavissa: <https://www.aclweb.org/anthology/K17-3009.pdf>
- Strömmer, Maiju 2017. *Mahdollisuuksien rajoissa: neksusanalyysi suomen kielen oppimisesta siivoustyössä*. Väitöskirja. Jyväskylän yliopisto. [Viitattu 16.7.2020.] Saatavissa: <https://jyx.jyu.fi/handle/123456789/56113>
- Sundqvist, P. – Sylvén, L.K. 2014. *Language-Related Computer Use: Focus on Young L2 English Learners in Sweden* \*. ReCALL (Cambridge, England) 26:1, s. 3–20. [Viitattu 15.4.2020.] Saatavissa: <https://doi.org/10.1017/S0958344013000232>
- Sundqvist, Pia 2009. *Extramural English Matters: Out-of-School English and Its Impact on Swedish Ninth Graders' Oral Proficiency and Vocabulary*. Karlstad University. [Viitattu 16.5.2020.] Saatavissa: <http://kau.diva-portal.org/smash/record.jsf?pid=diva2%3A275141&dswid=-1593>
- Sundqvist, Pia 2011. *A possible path to progress: Out-of-school English language learners in Sweden*. Teoksessa Benson, P. and Reinders, H. (eds.), *Beyond the language classroom*. Basingstoke: Palgrave Macmillan, s. 106–118. [Viitattu 16.5.2020.] Saatavissa: [https://doi.org/10.1057/9780230306790\\_9](https://doi.org/10.1057/9780230306790_9)
- Suni, Minna 2008. *Toista kieltä vuorovaikutuksessa: Kielellisten resurssien jakaminen toisen kielen oppimisen alkuvaiheessa*. Väitöskirja. Jyväskylän yliopisto. Viitattu [15.4.2020.] Saatavissa: <http://urn.fi/URN:ISBN:978-951-39-3209-1>
- Suni, Minna 2010. *Työssä opittua: työntekijän näkökulma ammatilliseen kieli- ja viestintätaitoonsa*. Teoksessa Mikel Garant – Mirja Kinnunen (toim.) AFinLA-e: Soveltavan kielitieteen tutkimuksia 2, s. 45–58. [Verkkolehti.] Jyväskylä: AFinLA. [Viitattu 15.4.2020.] Saatavissa: <https://journal.fi/afinla/article/view/3875>
- Sylvén, L. K., – Sundqvist P. 2012. *Gaming as Extramural English L2 Learning and L2 Proficiency Among Young Learners* \*. ReCALL (Cambridge, England) 24: 3, s. 302–321. [Viitattu 15.4.2020.] Saatavissa: <https://doi.org/10.1017/S095834401200016X>

- Sylvén, L.K. – Sundqvist, P. 2017. *Computer-Assisted Language Learning (CALL) in Extracurricular/extramural Contexts*. CALICO Journal 34:1: i–iv. [Viitattu 15.4.2020.] Saatavissa: <https://doi.org/10.1558/cj.31822>
- Te-palvelut. Työ- ja elinkeinotoimiston kotisivu [Viitattu 15.4.2020.] Saatavissa: [https://www.te-palvelut.fi/te/fi/tyonhakijalle/ammatinvalinta\\_koulutus/tyovoimakoulutus/index.html](https://www.te-palvelut.fi/te/fi/tyonhakijalle/ammatinvalinta_koulutus/tyovoimakoulutus/index.html)
- Tervola, Maija 2019. *Maahanmuuttajalääkärien suomen kielen taito ja kielitaitotarpeet: Erityisalan kielitaidon näkökulma lääkärin työhön*. Tohtorinväitöskirja. Tampereen yliopisto. [Viitattu 12.11.2020.] Saatavissa: <https://trepo.tuni.fi/handle/10024/116309>
- Tilastokeskus [Viitattu 12.11.2020.] Saatavissa: <https://www.stat.fi/tup/maahanmuutto/maahanmuuttajat-vaestossa.html>
- UDPipe [Viitattu 23.10.2019.] Saatavissa: <http://lindat.mff.cuni.cz/services/udpipe/>
- Valvira.fi [Viitattu 28.1.2019.] Saatavissa: [https://www.valvira.fi/terveydenhuolto/ammattioikeudet/hakemusohjeet/eu\\_eta-valtioissa\\_koulutetut/laillistettavat](https://www.valvira.fi/terveydenhuolto/ammattioikeudet/hakemusohjeet/eu_eta-valtioissa_koulutetut/laillistettavat)
- van Lier, Leo 1996. *Interaction in the language curriculum: Awareness, autonomy and authenticity*. Longman, London, New York.
- van Lier, Leo 2000. *From input to affordance: Social-interactive learning from an ecological perspective*. Teoksessa J. P. Lantolf (ed.), *Sociocultural theory and second language learning*. Oxford: Oxford University Press.
- van Lier, Leo 2004. *The Ecology and Semiotics of Language Learning: a Sociocultural Perspective*. Boston: Kluwer Academic Publishers.
- van Lier, Leo 2007. *Action-based Teaching, Autonomy and Identity*. International Journal of Innovation in Language Learning and Teaching, 1:1, s. 46–65. [Viitattu 1.6.2020.] Saatavissa: <https://doi.org/10.2167/illt42.0>
- van Lier, Leo 2010. *The ecology of language learning: Practice to theory, theory to practice*. Procedia Social and Behavioral Sciences 3, s. 2–6. [Viitattu 1.6.2020.] Saatavissa: <http://dx.doi.org/10.1016/j.sbspro.2010.07.005>
- Virtanen, A. 2011. *Käsityksiä kansainvälisesti rekrytoitujen hoitajien ammatillisesta kielitaidosta ja sen kehittämisestä. Mediakeskusteluiden ja asiantuntijan haastattelun analyysia*. Puhe ja kieli, 31: 4, s. 153–172. [Verkkolehti.] Helsingin yliopisto. [Viitattu 16.7.2020.] Saatavissa: <https://journal.fi/pk/article/view/4751>
- Virtanen, A. 2017 a. *Toimijuutta toisella kielellä: kansainvälisten sairaanhoitajaopiskelijoiden ammatillinen suomen kielen taito ja sen kehittyminen työharjoitteluissa*. Väitöskirja. Jyväskylän yliopisto. [Viitattu 16.7.2020.] Saatavissa: <http://urn.fi/URN:ISBN:978-951-39-7021-5>
- Virtanen, A. 2017 b. *Toimijuutta toisella kielellä. Kansainvälisten sairaanhoitajaopiskelijoiden ammatillinen suomen kielen taito ja sen kehittyminen työharjoitteluissa*. Virittäjä 121: 3 [Verkkolehti.] Kotikielen Seura, Helsinki. [Viitattu 18.8.2020.] Saatavissa: <https://doi.org/10.23982/vir.63695>
- Webb Stuart – Rodgers Michael P. H. 2009. *The Lexical Coverage of Movies Applied Linguistics*, 30: 3, s. 407–427, Saatavissa: <https://doi.org/10.1093/applin/amp010>
- Yliniemi, M. – Dufva, H. – Mäntylä, K. 2006. *"Jatkuvasti tulee puhuttua" -lasten käsityksiä kieliympäristön merkityksestä vieraan kielen oppimisessa*. Teoksessa R. Alanen, H. Dufva, & K. Mäntylä (Eds.), *Kielen päällä: Näkökulmia kieleen ja kielenkäyttöön* s. 77–96. Solki, Jyväskylä



## Liite 1 Taustatietolomake

### Tutkintokieli ja -taso:

Kieli	<input type="checkbox"/> Englanti
	<input type="checkbox"/> Espanja
	<input type="checkbox"/> Italia
	<input type="checkbox"/> Ranska
	<input type="checkbox"/> Ruotsi
	<input type="checkbox"/> Saame
	<input type="checkbox"/> Saksa
	<input type="checkbox"/> Suomi
	<input type="checkbox"/> Venäjä
Taso	<input type="checkbox"/> Perustaso
	<input type="checkbox"/> Keskitaso
	<input type="checkbox"/> Ylin taso

### Tasoarviot:

Yleistasoarvio	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Puheen ymmärtäminen	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Tekstin ymmärtäminen	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Puhuminen	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Kirjoittaminen	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Rakenne ja sanasto	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Hakuehtotyyppi:	<input checked="" type="radio"/> joko / tai						
	<input type="radio"/> sekä / että						

### Taustamuuttajat:

Ikä:	<input type="range"/>			
	12 - 80			
Sukupuoli	<input type="checkbox"/> Nainen	<input type="checkbox"/> Mies		
Äidinkieli	<input type="checkbox"/> Suomi	<input type="checkbox"/> Ruotsi	<input type="checkbox"/> Venäjä	<input type="checkbox"/> Albania
	<input type="checkbox"/> Arabia	<input type="checkbox"/> Dari	<input type="checkbox"/> Englanti	<input type="checkbox"/> Espanja
	<input type="checkbox"/> Italia	<input type="checkbox"/> Kiina	<input type="checkbox"/> Kurdi	<input type="checkbox"/> Ranska
	<input type="checkbox"/> Saame	<input type="checkbox"/> Saksa	<input type="checkbox"/> Somali	<input type="checkbox"/> Turkki
	<input type="checkbox"/> Viro	<input type="checkbox"/> Muu		
Peruskoulutus	<input type="checkbox"/> Peruskoulu / kansakoulu			
	<input type="checkbox"/> Lukio			
	<input type="checkbox"/> Ammatillinen oppilaitos			
	<input type="checkbox"/> Ammattikorkeakoulu			
	<input type="checkbox"/> Yliopisto			
	<input type="checkbox"/> Muu			
Sosioekonominen asema	<input type="checkbox"/> Yrittäjä			
	<input type="checkbox"/> Ylempi toimihenkilö			
	<input type="checkbox"/> Alempi toimihenkilö			
	<input type="checkbox"/> Työntekijä			
	<input type="checkbox"/> Opiskelija			
	<input type="checkbox"/> Työvoimakoulutuksen opiskelija			
	<input type="checkbox"/> Eläkeläinen			
	<input type="checkbox"/> Työtön			
	<input type="checkbox"/> Muu			

Toimiala	<input type="checkbox"/> Maa- ja metsätalous
	<input type="checkbox"/> Teollisuus
	<input type="checkbox"/> Rakennustoiminta
	<input type="checkbox"/> Kauppa, majoitus- ja ravitsemistoiminta
	<input type="checkbox"/> Liikenne
	<input type="checkbox"/> Rahoitus, vakuutus ja liike-elämää palveleva toiminta
	<input type="checkbox"/> Maanpuolustus
	<input type="checkbox"/> Julkiset ja muut palvelut
	<input type="checkbox"/> Muu

Missä ja kuinka kauan olet opiskellut testattavaa kieltä?

Hakuehtotyyppi: ?	<input checked="" type="radio"/> joko / tai <input type="radio"/> sekä / että
-------------------	--

Peruskoulu / kansakoulu	<input type="checkbox"/> Alle vuoden	<input type="checkbox"/> 1-2 vuotta	<input type="checkbox"/> 3-5 vuotta	<input type="checkbox"/> 6-9 vuotta
	<input type="checkbox"/> 10 vuotta tai yli	<input type="checkbox"/> En lainkaan		

Lukio	<input type="checkbox"/> Alle vuoden	<input type="checkbox"/> 1-2 vuotta	<input type="checkbox"/> 3-5 vuotta	<input type="checkbox"/> 6-9 vuotta
	<input type="checkbox"/> 10 vuotta tai yli	<input type="checkbox"/> En lainkaan		

Ammatillinen oppilaitos	<input type="checkbox"/> Alle vuoden	<input type="checkbox"/> 1-2 vuotta	<input type="checkbox"/> 3-5 vuotta	<input type="checkbox"/> 6-9 vuotta
	<input type="checkbox"/> 10 vuotta tai yli	<input type="checkbox"/> En lainkaan		

Ammattikorkeakoulu	<input type="checkbox"/> Alle vuoden	<input type="checkbox"/> 1-2 vuotta	<input type="checkbox"/> 3-5 vuotta	<input type="checkbox"/> 6-9 vuotta
	<input type="checkbox"/> 10 vuotta tai yli	<input type="checkbox"/> En lainkaan		

Yliopisto	<input type="checkbox"/> Alle vuoden	<input type="checkbox"/> 1-2 vuotta	<input type="checkbox"/> 3-5 vuotta	<input type="checkbox"/> 6-9 vuotta
	<input type="checkbox"/> 10 vuotta tai yli	<input type="checkbox"/> En lainkaan		

Kansan- / kansalais- / työväenopisto	<input type="checkbox"/> Alle vuoden	<input type="checkbox"/> 1-2 vuotta	<input type="checkbox"/> 3-5 vuotta	<input type="checkbox"/> 6-9 vuotta
	<input type="checkbox"/> 10 vuotta tai yli	<input type="checkbox"/> En lainkaan		

Itsenäinen opiskelu	<input type="checkbox"/> Alle vuoden	<input type="checkbox"/> 1-2 vuotta	<input type="checkbox"/> 3-5 vuotta	<input type="checkbox"/> 6-9 vuotta
	<input type="checkbox"/> 10 vuotta tai yli	<input type="checkbox"/> En lainkaan		

Työnantajan järjestämät kurssit	<input type="checkbox"/> Alle vuoden	<input type="checkbox"/> 1-2 vuotta	<input type="checkbox"/> 3-5 vuotta	<input type="checkbox"/> 6-9 vuotta
	<input type="checkbox"/> 10 vuotta tai yli	<input type="checkbox"/> En lainkaan		

Työvoimakoulutus	<input type="checkbox"/> Alle vuoden	<input type="checkbox"/> 1-2 vuotta	<input type="checkbox"/> 3-5 vuotta	<input type="checkbox"/> 6-9 vuotta
	<input type="checkbox"/> 10 vuotta tai yli	<input type="checkbox"/> En lainkaan		

Ulkomailta	<input type="checkbox"/> Alle vuoden	<input type="checkbox"/> 1-2 vuotta	<input type="checkbox"/> 3-5 vuotta	<input type="checkbox"/> 6-9 vuotta
	<input type="checkbox"/> 10 vuotta tai yli	<input type="checkbox"/> En lainkaan		

Muu	<input type="checkbox"/> Alle vuoden	<input type="checkbox"/> 1-2 vuotta	<input type="checkbox"/> 3-5 vuotta	<input type="checkbox"/> 6-9 vuotta
	<input type="checkbox"/> 10 vuotta tai yli	<input type="checkbox"/> En lainkaan		

## Missä ja kuinka usein käytät testattavaa kieltä?

Hakuehtotyyppi: ☒ joko / tai  
☐ sekä / että

Koti	<input type="checkbox"/> Lähes päivittäin	<input type="checkbox"/> Kerran pari viikossa	<input type="checkbox"/> Kerran kuussa tai harvemmin	<input type="checkbox"/> En lainkaan
Työ	<input type="checkbox"/> Lähes päivittäin	<input type="checkbox"/> Kerran pari viikossa	<input type="checkbox"/> Kerran kuussa tai harvemmin	<input type="checkbox"/> En lainkaan
Vapaa-aika	<input type="checkbox"/> Lähes päivittäin	<input type="checkbox"/> Kerran pari viikossa	<input type="checkbox"/> Kerran kuussa tai harvemmin	<input type="checkbox"/> En lainkaan
Opiskelu	<input type="checkbox"/> Lähes päivittäin	<input type="checkbox"/> Kerran pari viikossa	<input type="checkbox"/> Kerran kuussa tai harvemmin	<input type="checkbox"/> En lainkaan

## Mistä sait tietoa Yleisistä kielitutkinnoista?

- ☐ Työnantajalta
- ☐ Opettajalta
- ☐ Tuttavalta
- ☐ Internetistä
- ☐ Omasta oppilaitoksesta
- ☐ Tutkintoa järjestävästä oppilaitoksesta
- ☐ Työvoimatoimistosta
- ☐ Julkisten tilojen ilmoitustaululta
- ☐ Testiesitteestä
- ☐ Lehti-ilmoituksesta
- ☐ Muualta

## Miten valitsit tutkintotason?

- ☐ Työnantajan vaatimuksesta
- ☐ Opettajan suosituksesta
- ☐ Oman arvion perusteella
- ☐ Kurssin järjestäjän avustuksella
- ☐ Tutkinto sisältyy kielikurssiin
- ☐ Tutustumalla testiesitteeseen
- ☐ Lukemalla tasokuvauksia
- ☐ En osaa sanoa
- ☐ Muuten

## Mihin tulet käyttämään tutkintotodistusta?



- ☐ Työnhakuun
- ☐ Opiskeluun
- ☐ Saadakseni palautetta kielitaidostani
- ☐ Osoittaakseni kielitaitoni työnantajalle
- ☐ Saadakseni tutkintokielen osaamista vaativia työtehtäviä
- ☐ Hakeakseni kielilisää palkkaan
- ☐ Hakeakseni Suomen kansalaisuutta
- ☐ Muuhun

## Kirjoittamisen suoritukset

- ☐ Rajaa hakutulos sisältämään vain ne tapaukset, joilta on tallennettuna kirjoittamisen suoritus

<b>Tekstilaji</b>	Valitse...
<b>Tehtävätyyppi</b>	Valitse...
<b>Sanahaku</b>	Muodollinen kirje Epämuodollinen kirje Muodollinen viesti Epämuodollinen viesti Palaute Mielipideteksti Ilmoitus Raportti Lausunto Hakemus Tiedote Ohje Muistio

## Puhumisen suoritukset

- ☐ Rajaa hakutulos sisältämään vain ne tapaukset, joilta on tallennettuna puhumisen suoritus

## Kirjoittamisen suoritukset

- ☐ Rajaa hakutulos sisältämään vain ne tapaukset, joilta on tallennettuna kirjoittamisen suoritus

<b>Tekstilaji</b>	Valitse...
<b>Tehtävätyyppi</b>	Valitse...
<b>Sanahaku</b>	Ohjattu Otsikkopohjainen

## Puhumisen suoritukset

- ☐ Rajaa hakutulos sisältämään vain ne tapaukset, joilta on tallennettuna puhumisen suoritus

Lähde: YKI-korpus. Jyväskylän yliopiston Soveltavan kielentutkimuksen keskus.

## Liite 2 Taitotasokuvaukset

Yleisten kielitutkintojen taitotasot			EVK:n* asteikko
Ylin taso	6	Ymmärtää vaikeuksitta kaikenlaista puhuttua ja kirjoitettua kieltä. Ilmausten hienot sävyerot tuottavat enää harvoin vaikeuksia. Puhuu ja kirjoittaa erittäin sujuvasti tilanteeseen sopivalla tyyllillä ja pystyy ilmaisemaan hienojakin merkitysvivahteita. Kieliopin ja sanaston hallinta on varmaa lähes kaikissa tilanteissa, ja pienetkin puutteet ovat harvinaisia.	C 2
	5	Ymmärtää kaikenlaista normaalitempoista puhetta, vaikka joskus ymmärtäminen voi vaatia ponnistelua, jos aihepiiri tai asian käsittely on vierasta. Ymmärtää rakenteellisesti ja kielellisesti monimutkaisia tekstejä ja oman aikamme kirjallisuutta. Puhuu ja kirjoittaa selkeästi ja sujuvasti erilaisista aiheista, mutta harvinaisten sanojen ja monimutkaisten lauserakenteiden käyttö saattaa kuitenkin tuottaa vaikeuksia.	C 1
Keskitaso	4	Ymmärtää normaalitempoista puhetta yleisistä aihepiireistä, mutta joitakin yksityiskohtia saattaa jäädä ymmärtämättä. Nopea puhekieli ja murteellisuudet tuottavat kuitenkin vaikeuksia. Ymmärtää vaivatta yleisiä aihepiirejä käsitteleviä tekstejä, joskin muutamien tekstin välittämät sävyerot voivat jäädä epäselviksi. Selviää melko hyvin erilaisissa sekä virallisissa että epävirallisissa puhetilanteissa. Pystyy kirjoittamaan sekä yksityisiä että puolivirallisia tekstejä ja esittämään ajatuksiaan yhtenäisinä kokonaisuuksina.	B 2
	3	Ymmärtää pidempää yhtäjaksoista puhetta ja keskeisen ajatuksen monista televisio- ja radio-ohjelmista, jos aihepiiri on tuttu ja puhetempo normaali. Ymmärtää tavallisia jokapäiväisen elämän tekstejä, mutta vaativammat tekstit, joiden aihepiiri on vieras, saattavat tuottaa vaikeuksia. Selviää tavallisimmissa käytännön puhetilanteissa ja pystyy kirjoittamaan yksinkertaista, yhtenäistä tekstiä tavallisista aiheista, vaikkakin kieliopilliset ja sanastolliset puutteet toisinaan vaikeuttavat ymmärtämistä.	B 1
Perustaso	2	Ymmärtää selkeää ja yksinkertaistettua puhetta, joka käsittelee jokapäiväisiä ja tuttuja asioita. Ymmärtää helposti lyhyitä, yksinkertaisia tekstejä ja saa selville pääasiat jokapäiväisen elämän aihepiirejä käsittelevistä teksteistä. Selviää rutiininomaisissa yksinkertaista tiedonvaihtoa vaativissa puhetilanteissa, vaikka ääntäminen tai kielen hallinta voi olla vielä puutteellista. Pystyy kirjoittamaan suppeita, yksinkertaisia tekstejä jokapäiväisistä asioista, mutta teksti voi olla hajanaista.	A 2
	1	Ymmärtää hitaasta ja selkeästä puheesta yksinkertaisia perustason ilmauksia, jotka liittyvät suoraan omaan elämään tai jotka koskevat välitöntä konkreettista ympäristöä. Pystyy ymmärtämään joitakin asioita helppotajuisista ja lyhyistä teksteistä. Selviää kaikkein yksinkertaisimmissa puhetilanteissa, mutta puhe on hidasta ja hyvin katkonaista ja ääntämisessä ja/tai kielen hallinnassa voi olla puutteita. Pystyy kirjoittamaan erittäin lyhyitä tekstejä, joissa on lukuisia kielellisiä puutteita.	A 1

Yleisten kielitutkintojen taitotasosteikko on linkitetty empiirisesti eurooppalaisen viitekehyksen (EVK) asteikkoon.

Lähde: *Yleisten kielitutkintojen perusteet 2011*. Määräys 24/011/2011. Opetushallitus.  
[https://www.oph.fi/sites/default/files/documents/169733\\_taitotasokuvaukset\\_cefr.pdf](https://www.oph.fi/sites/default/files/documents/169733_taitotasokuvaukset_cefr.pdf)